

‘Fake News’: Legal and Psychological Challenges in the Regulation of Social Media

Bradley S. Foster, J.D., LL.M.

100145506

Submitted for qualification as Doctor of Philosophy in Law

University of East Anglia, School of Law

12 May 2023

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there-from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Abstract

The role of social media news use in recent elections in the United States and the United Kingdom has prompted the introduction of various policy proposals and interventions by legislatures and social media companies. However, the scale of global social media enterprises, ranging from hundreds of millions to billions of users worldwide, has produced voluntary and co-regulatory approaches that rely on the technical capabilities of the social media platforms to implement policymakers' objectives. YouTube, Twitter, and Facebook use a combination of rules or guidelines and algorithmic automation to make decisions about user-generated content at scale. A substantial part of the data social media platforms rely on to moderate content comes from the platforms' users, especially through content flagging tools that allow users to report instances of fake news.

Therefore, a key consideration in the evaluation of policy proposals to regulate social media should be the quality of those user-data inputs upon which regulation may rely for its implementation. I designed my research to examine the relationship of political group identification to making judgments about the political news stories users encounter on social media. My study assessed American participants' level of identification with political parties ahead of the 2020 U.S. election and then tasked participants with evaluating genuine and fictitious political news and rating their likelihood to report the stories as false. I found that the more highly identified participants in each group tended to report genuine stories that denigrated their own side as false.

I further contend that proposals to regulate social media content should only proceed with a robust understanding of the trade-offs inherent in technical solutions to human circumstances. In addressing news content on social media, we should be wary about institutionalizing algorithms as the arbiters of truth—especially where they mimic and reinforce our group psychological tendencies.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate 'take down' action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

List of Contents

Introduction.....	6
Part One: The Problem of ‘Fake News’ on Social Media.....	8
Social Media News Use	8
The Fake News Problem	14
Fake News Studies.....	30
Part Two: Approaching Fake News as a Media System Problem.....	36
Social Media and Media Systems.....	37
Platform Policies and Technical Capabilities	64
Part Three: The Social Psychology of Decision Making.....	87
Introduction to Law and Psychology	87
Principles of Social Psychology.....	89
The Social Psychology of Political News and Social Media.....	101
Constraints on Content Moderation Policy Effectiveness	106
Part Four: Experiment Design and Results	113
Methodology	113
Findings	129
Implications	143
Part Five: Analysis and Conclusions about Proposals to Regulate Social Media.....	147
Public Policy and the Public Sphere.....	150
Analysing Allocation Decisions	175
Conclusions	205
Appendix A	211
Test Instrument.....	211
Appendix B	242
Research Ethics Application.....	242
Appendix C.....	255
Additional Measurements.....	255
Supporting Analyses.....	260
Future Research Supplementals.....	263
Appendix D	273
20 th Century American Jurisprudence	273
Bibliography	293

List of Figures and Tables

Figure 1 – Google Trends: ‘labour antisemitism’	22
Table 1 – Participant Demographics	128
Table 2 – Comparisons Between Participant Groups by Group Identification Scores	130
Table 3 – Comparisons Between Participant Groups by Sharing and Flagging Scales	131
Table 4 – Republican Participant Sharing and Flagging Comparison by News Item Category	132
Table 5 – Democrat Participant Sharing and Flagging Comparison by News Item Category	132
Table 6 – Comparison Between Groups by Relevant Decision Scores	133
Table 7 – Kendall Rank Correlation Table	134
Table 8 – Group Separated Kendall Rank Correlation Table	135
Table 9 – Regression analysis predicting willingness to share fake outgroup stories from demographic factors and group identification scores, interaction	136
Table 10 – Group separated regression analyses predicting willingness to share fake outgroup stories from demographic factors and group identification scores ...	137
Table 11 – Regression analysis predicting willingness to flag genuine ingroup stories as fake from demographic factors and group identification scores, interaction ...	138
Table 12 – Group separated regression analyses predicting willingness to flag genuine ingroup stories as fake from demographic factors and group identification scores	139

*An additional List of Tables is provided in **Appendix C** for the tables therein.*

Acknowledgments

My sincerest thanks to my family back home, without whose support these past years of solitary work would have been impossible. Your example, guidance, and love mean everything to me. My thanks also to my UEA School of Law PGR cohorts for all our efforts to sustain and encourage each other through particularly hard times. It has been a privilege to know you all. And thanks once more to David, Karen, and Charlie for sticking by me over the long evolution of my thesis. Your patience, curiosity, and enthusiasm have been invaluable.

Introduction

Over the course of my research into how the law might address the issue of fake political news stories on social media, I noticed a problem. How fake news is studied and the proposals for doing something about it tend to focus on particular aspects of online content itself which are assumed to be dispositive of truth or authenticity. That makes sense from the perspective of the social media platforms, who rely on various means of identifying and acting upon the vast amounts of news their users share. The operation of their content moderation systems (automated computer processes for managing posts, comments, videos, etc.) need discrete inputs about user-generated content to generate specific outcomes, such as blocking or demonetizing a post. Lawmakers and regulators likewise aim for specific outcomes regarding objectionable content like fake news—especially when it bears on politics and the elections process. The operation of the platforms' content moderation systems is thus integral to achieving such policy objectives where laws seek to hold the platforms to account for content that may mislead the public or interfere with elections.

However, studies of fake news and related policy interventions vary widely in approach—often due to arbitrary or inapt definitions of 'fake news' and the difficulty of interpreting observations about content relative to the platforms' inscrutable proprietary processes. Accordingly, I determined that an experimental research approach could avoid these obstacles and provide novel insights into an important input for content moderation on social media platforms. These systems incorporate data generated by users—inputs such as flags, reports, likes, and shares—that tell the platforms about the content the users encounter. User feedback (and even expert annotation), especially about political news, involves an exercise of human judgment that is essentially social in nature. Drawing on principles of social psychology, I reasoned that users' decision making about the authenticity of political news would be related to their political affiliation.

Identifying the relatedness of political group membership and news judgments in turn casts new light on the reliability of the user data inputs leveraged by the platforms. That may help explain observed phenomena like false positives, where content moderation systems take action against genuine political news identified as fake. False positives in particular bear on whether the policy aims of lawmakers can effectively combat fake news while respecting human rights to free expression and information. Where regulatory proposals rely on the platforms' moderation systems to implement policy aims, efforts to improve the civic function of social media may be stymied as a result of moderation not working as intended. Recognizing this potential helps us better understand the issue and discern between alternative approaches to promoting public discourse over social media.

- In Part One of this Thesis, I survey some of the ways social media is used for political news production and consumption and how fake news is being studied. This survey situates the political and electoral consequences of social media news use that have made this issue acute—and points out challenges encountered in its examination.
- In Part Two, I survey past social and legal responses to disruptive media technology advances through the lens of media systems. This lens highlights how those responses relate to media system structure and function. Looking at differences and similarities in structure and function between earlier news media and social media in turn suggests where future policy responses may be more or less well aligned with social media than with past systems. In particular, automated content moderation demands closer scrutiny as an element that sets apart leading social media platforms like Facebook, YouTube, and Twitter.
- The social psychological principles I review in Part Three help us understand more about the data ‘inputs’ into automated moderation systems provided to these platforms by social media users concerning political news. Looking at the social dimension of decision making opens up a rich account of the mental processes involved when online news bears on a particular political group.
- The experiment I detail in Part Four allowed me to test how social identity—that part of our sense of self linked to the groups to which we belong—relates to the kinds of decisions we make about political news on social media. My findings show that decisions to share political news stories or flag them as fake are associated with more besides a mere desire for accuracy or fairness. Understanding decisions like these that factor into automated content moderation systems helps us evaluate both the status quo and policy proposals that rely on such systems.
- In Part Five, I review and critique current and developing social media policies in Britain and the United States in light of my analysis and findings. I further consider additional alternatives for improving the civic function of social media that look beyond reliance on automated moderation. I conclude by evaluating the trade-offs in values, interests, and results represented by some of the choices available to us about accountability for fake news.

Part One: The Problem of ‘Fake News’ on Social Media

This Part introduces the problem of ‘fake news’ through three key discussions. The first discussion covers the ways social media is used for news production and consumption, including how social media figures alongside traditional news use and political activities. The second discussion explores the idea of ‘fake news’—the notion that information shared on social media with the form or characteristics of news will be received by audiences as trustworthy, particularly with regard to political issues and elections. The third discussion looks at some of the studies that aim to provide a social science context for describing fake news and how or why it may spread over social media. The aim of these discussions is to situate the problem of fake news relative to specific stakeholders and explore how certain uses of social media are described as problematic. The next Part of this Thesis will discuss how we might think about the social media platforms and investigates their key features.

Social Media News Use

In 2018, the president of Nigeria released a video statement to his nearly two million Twitter followers denying rumours being spread on social media, including YouTube and Facebook videos garnering thousands of views, that he had died and been replaced by a doppelganger.¹ The curious incident arose amid increasing use of social media by people around the world as an important source of news. In the United States, one-in-five adults reported in 2018 that they often got their news from social media, surpassing the number who similarly relied on print newspapers.² While still the most popular news source, television has declined in popularity since 2016 in the U.S.³ Likewise in the United Kingdom, television is the news source U.K internet users generally turn to first, but social media has become the most popular source among U.K. internet users who want news providing alternative viewpoints.⁴ News as an essential idea predates printing and even writing, and in textual form has been dated to Roman times as the *Acta Diurna*, while the “origins of the modern newspaper are usually attributed to the newsletters that circulated in the latter sixteenth century, especially the Venetian *corantos* or accounts of current events . . . [which] mainly disseminated factual information about recent political and military events that were

¹ Guardian staff and agencies, “‘It’s the real me’: Nigerian president denies dying and being replaced by clone” (*The Guardian*, 03 December 2018) <<https://www.theguardian.com/world/2018/dec/03/its-real-me-nigerian-president-denies-dying-and-being-replaced-by-clone>> accessed 28 July 2022

² Elisa Shearer, ‘Social media outpaces print newspapers in the U.S. as a news source’ (*Pew Research Center*, 10 December 2018) <<https://www.pewresearch.org/fact-tank/2018/12/10/social-media-outpaces-print-newspapers-in-the-u-s-as-a-news-source/>> accessed 28 July 2022

³ *ibid*

⁴ Ofcom, *Adults’ Media Use and Attitudes Report 2018* (25 April 2018) 72-73

of relevance to trade and commerce.”⁵ Social media news use sits at the frontier of a vast continent of communications technologies that have been charted, enclosed, managed, and developed over a considerable span of history.

While social media today is widely understood to offer users a megaphone for easily and affordably sharing information farther and more widely than ever, earlier technologies like the printing press and telegraph were also deeply disruptive in the centuries of their development. Manuscript copying in the Fourteenth Century took months and months to complete, suffered due to illegibility and error in the originals, and effectively bottlenecked how readily texts could be retrieved or referenced.⁶ But the falling price of paper in Europe (thanks to the advent of water-powered paper mills) and Gutenberg’s development of the moveable typeface printing press in the 1450s⁷ greatly increased the productivity of the workshops producing manuscript books.⁸ After the invention of Gutenberg’s printing press, one of Gutenberg’s financial backers carried twelve of the printer’s Bibles with him to Paris, only to be chased off after the long journey by the Parisian book trade guilds on suspicion the identical copies must have been produced with the help of the devil.⁹ Social media likewise has increased the productivity of news producers, albeit by reducing the time needed to accurately reproduce a text from minutes to moments—rather than from months to minutes—and perhaps with slightly less suspicion of supernatural assistance.

Since the time of Gutenberg, communications technology has also profoundly altered information’s speed of transmission and cost of distribution—two hallmarks of today’s information age. By the Fifteenth Century, courier services were regularly employed to relay information by “the Roman Curia and the royal houses of England, Aragon, the republic of Venice and the university of Paris.”¹⁰ Posting messengers who could rely on regularly stationed fresh horses were able to cover an average of ninety miles a day, “more than twice as much as an ordinary rider,” and were employed by the Medici bankers to keep in touch with their branch managers and representatives across Europe.¹¹ The horse-drawn stagecoach remained the standard for American postal delivery over long distances into the middle of the Nineteenth Century: the overland route from St. Louis, Missouri to San Francisco, California (around 2,800 miles) took 21 ½ days by stagecoach.¹² The famed “Pony Express” used the same technique as the Medicis hundreds of years before, and riders

⁵ Denis McQuail, *Media Accountability and Freedom of Publication* (OUP 2007) 32-33

⁶ James Burke, *The Day the Universe Changed* (rev edn, Little, Brown 1995) 104-07

⁷ *ibid* 109-12

⁸ McQuail (n 5) 28

⁹ Burke (n 6) 113

¹⁰ *ibid* 96

¹¹ *ibid*

¹² U.S. Postal Service, *The United States Postal Service: An American History* (Washington, D.C. 2020) 15

regularly changing horses could manage about 75 to 100 miles a day along a route to California from St. Joseph, Missouri—"the westernmost point [then] reached by railroad and telegraph"—that cut the stagecoach time in half.¹³ The further expansion of the railroads westward, however, along with their accompanying telegraph lines, fundamentally shifted the economics of information transmission: "On October 26, 1861 – two days after the transcontinental telegraph line was completed – the Pony Express officially ended."¹⁴ Electronic communications technologies in use today, including the Internet, continue to capitalize on the dramatic reduction in time and cost for information distribution accomplished by the telegraph.

Social media can certainly seem like the Wild West in comparison to earlier modes of communication used for sharing news. News professionals have struggled with declining trust in traditional media, particularly in recent years.¹⁵ That phenomenon is not unique to American political circumstances, and distrust in traditional media has not apparently translated into any greater trust of alternative media.¹⁶ Researchers have instead found that people in countries from the U.K. to Brazil to Japan are increasingly avoiding the news altogether.¹⁷ The user experience of online news offers little to suggest a competitive advantage over traditional media: The typical news website is replete with clickbait headlines, cookie preference pop-ups, auto-play video advertisements, insistent subscription offers, and precious screen space devoted to stories other than the one you wanted to read.¹⁸ Gil de Zuniga, Jung, and Valenzuela argue that the motivations for using online services are the key to understanding the place of social media in daily life—activities like chatting, commenting, posting links, and sharing news stories as functions of social dynamics like community, status, and identity.¹⁹ Social media news-use in particular engages a range of mental processes related to self-perception and motivation.²⁰ One of the primary objectives of this Thesis is to explore the role group identity plays in shaping decisions made about political news encountered on social media.

¹³ U.S. Postal Service (n 12) 16-17

¹⁴ *ibid* 17

¹⁵ Felix Salmon, 'Trust in media hits new crisis low' (*Axios*, 21 January 2021)

<<https://www.axios.com/2021/01/21/media-trust-crisis>> accessed 28 July 2022

¹⁶ *ibid*

¹⁷ Laura Hazard Owen, 'As traditional news use declines, online news isn't making up the gap' (*NiemanLab*, 14 June 2022) <<https://www.niemanlab.org/2022/06/as-traditional-news-use-declines-online-news-isnt-making-up-the-gap/>> accessed 28 July 2022

¹⁸ David Roth, 'The infinite scroll' (*Columbia Journalism Review*, 04 March 2020)

<https://www.cjr.org/first_person/the-infinite-scroll.php> accessed 28 July 2022

¹⁹ Homero Gil de Zuniga, Nakwon Jung, Sebastian Valenzuela, 'Social Media Use for News and Individuals' Social Capital, Civic Engagement and Political Participation' (2012) 17 *J Computer-Mediated Communication* 319, 321-22

²⁰ Nicolas M Anspach, Jay T Jennings, and Kevin Arceneaux, 'A little bit of knowledge: Facebook's News Feed and self-perceptions of knowledge' (January-March 2019) *Research and Politics* 1, 2-3

Early scholarship about the civic role of news dissemination and discussion over the Internet ranged from the optimistic to the uncertain. Balkin expected that the “digital age” would provide “a technological infrastructure that greatly expands the possibilities for individual participation in the growth and spread of culture and thus greatly expands the possibilities for the realization of a truly democratic culture.”²¹ In the same period, McNicholas and Ward remarked on a more modest pace of development and the replication of traditional political activities online.²² But the interactivity and general utility of social media platforms like Facebook did lend themselves to adoption by users for engagement in public affairs, particularly by time of the 2008 U.S. presidential election.²³ Over the last eight years, Facebook and YouTube have remained the leading platforms for social media news consumers in major international media markets.²⁴ News producers, particularly independent outlets, have also looked to Facebook and other social media as vital avenues for building audiences.²⁵

The high expectations set forth over the past decade about the potential impact of social media on civic engagement suggest why social media continues to figure prominently in discussions of elections today. The historical roots of such expectations are clearly recognizable—the printing press and the easy production of the broadsheet helped new arrangements of social and political relations to emerge in centuries past: rulers could promulgate ordinances and proclamations, while dissidents like Martin Luther could disseminate grievances and petitions.²⁶ Social media use has been put forward as a means of increasing government transparency and fighting corruption,²⁷ as well as facilitating public engagement and provision of services.²⁸ Social media tools have been lauded for their democratic and revolutionary potential as well, figuring into the coordination of protest activities and foreign policy strategies.²⁹ Blogging and other forms of participatory media have not always been welcomed by various governments, most notably in China, where the

²¹ Jack M Balkin, ‘Digital speech and democratic culture: A theory of freedom of expression for the information society’ (2004) 79 NYU L Rev 1, 5

²² Anthony McNicholas and David Ward, ‘United Kingdom: New Media’ in Bernd-Peter Lange and David Ward (eds), *The Media and Elections: A Handbook and Comparative Study* (Taylor and Francis 2004) 159

²³ Gil de Zuniga, Jung, and Valenzuela (n 19)

²⁴ Hazard Owen (n 17)

²⁵ Matt Taibbi, ‘Who Will Fix Facebook?’ (*Rolling Stone*, 26 November 2018)

<<https://www.rollingstone.com/politics/politics-features/who-will-fix-facebook-759916/>> accessed 28 July 2022

²⁶ Burke (n 6) 117-18

²⁷ John C Bertot, Paul T Jaeger, and Justin M Grimes, ‘Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies’ (2010) 27 Government Information Quarterly 264

²⁸ John C Bertot, Paul T Jaeger, and Derek Hansen, ‘The impact of policies on government social media usage: Issues, challenges, and recommendations’ (2012) 29 Government Information Quarterly 30

²⁹ Clay Shirky, ‘The Political Power of Social Media’ (*Foreign Affairs*, January/February 2011)

<<https://www.foreignaffairs.com/political-power-social-media>> accessed 28 July 2022

government has implemented extensive content filtering systems to manage its citizens' use of the Internet, particularly with regard to political organization.³⁰

Participants in mass political events like the 2011 Tahrir Square protests in Egypt have reported social media served as important sources of information and organization, especially among protestors who skewed younger and more educated and among women.³¹ Online discussion afforded increased opportunities for dissident political opinion to circulate publicly, particularly after the introduction of Facebook in Arabic in 2009.³² Online social networking and smartphones have been recognized as important developments in U.S. protest movements as well.³³ The “#Occupy” protests in late 2011 exhibited similar “decentralized coordination” among discrete communities and actors as did previous direct action campaigns like the 1999 WTO protests in Seattle, for example, but relied extensively on social media platforms like Twitter for mobilization efforts rather than more intensive communications among participants via e-mail and mailing lists as used in the past.³⁴

In addition to their potential to support traditional protests, social media platforms have also been discussed as the sites of new forms of protest that might sidestep laws constraining comparable expressive activities in physical space³⁵: an unlimited number of protestors can ‘sit-in’ on a Twitter hashtag without disrupting public order, for instance. The notion of ‘cyberspace’ as a distinct place informs the discourses of some Internet rights activists,³⁶ but also lends itself to a comparative discussion of the role of place in traditional free speech jurisprudence, particularly with regard to its development around protest activities, as has been recognized by the U.N. Human Rights Committee.³⁷ Place plays a vital role in sustaining autonomy and collective self-determination, and its erosion impacts those values accordingly.³⁸ If social media platforms do essentially provide a public forum for political expression, then lessons learned from a much deeper historical background may be applied in the balancing of interests in free speech and public safety online. But while social

³⁰ Rebecca MacKinnon, ‘Flatter world and thicker walls? Blogs, censorship and civic discourse in China’ (2008) 134 Public Choice 31

³¹ Zeynep Tufekci and Christopher Wilson, ‘Social Media and the Decision to Participate in Political Protest: Observations from Tahrir Square’ (2012) 62 J Communications 363, 368-69

³² *ibid* 364-66 (situating social media alongside satellite television like Al-Jazeera in broadening the Arabic-language public sphere)

³³ Jeffrey S Juris, ‘Reflections on #Occupy Everywhere: Social media, public space, and emerging logics of aggregation’ (2012) 39(2) American Ethnologist 259

³⁴ *ibid* 266-67

³⁵ Christina E Wells, ‘The Promise and Peril of Protesting in the Internet Era’ (2012) University of Missouri Legal Studies Research Paper Series 2012-23, 10-12

³⁶ *e.g.*, John Perry Barlow, ‘A Declaration of the Independence of Cyberspace’ (*Electronic Frontier Foundation*, 08 February 1996) <<https://www.eff.org/cyberspace-independence>> accessed 28 July 2022

³⁷ United Nations Human Rights Committee, ‘General comment No. 37 (2020) on the right of peaceful assembly (article 21)’ (17 September 2020) CCPR/C/GC/37, para. 10

³⁸ Thomas P Crocker, ‘Displacing Dissent: The Role of “Place” in First Amendment Jurisprudence’ (2006-07) 75 Fordham L Rev 2587, 2601-08

media platforms often hold themselves out as servants of “the public conversation,”³⁹ and operate by the grace of public infrastructure investment and substantial legal protection,⁴⁰ Facebook, YouTube, and Twitter nevertheless tend to assert their own prerogatives when the public sphere pushes against their commercial goals.⁴¹

Despite the power of governments over access to the Internet, the greatest control over the political reach of social media is arguably exercised by the platforms themselves.⁴² Facebook’s prominent role in public life around the globe—with its nearly two billion daily active users⁴³—has necessitated the company undertake assessments of the impact of its platform policies on human rights⁴⁴ in countries like Indonesia and Sri Lanka.⁴⁵ Changes made by Facebook to its News Feed algorithm in 2018, intended to address misinformation by prioritizing local publishers and back-and-forth discussions, reportedly played a role in the rapid rise in France of the online *groupes colère* that launched the offline Yellow Vests protest movement.⁴⁶ An early study by Conover and others highlighted how certain functionalities of social media platforms like Twitter’s ‘retweet’ function tended to circulate political content within ideologically similar networks, while other functions like hashtags and ‘mentions’ had greater potential to reach ideologically opposed user networks.⁴⁷ While researchers have reached different conclusions over the past decade whether social media platforms produce a certain ‘echo chamber’ effect, where users see more news they agree

³⁹ Twitter, ‘About: Our Purpose’ (2022) <<https://about.twitter.com/en/who-we-are/our-company>> accessed 28 July 2022

⁴⁰ This notion is discussed further in later Parts of this Thesis

⁴¹ Arne Hintz, ‘Restricting digital sites of dissent: commercial social media and free expression’ (2016) 13 Critical Discourse Studies 325

⁴² Lawrence Lessig, *Code: Version 2.0* (Basic Books/Perseus 2006) 235-36 (pointing out the constraint and support of free expression embodied by the digital ‘architecture’ of online platforms)

⁴³ Facebook, ‘Facebook Reports Fourth Quarter and Full Year 2020 Results’ (27 January 2021) <<https://investor.fb.com/investor-news/press-release-details/2021/Facebook-Reports-Fourth-Quarter-and-Full-Year-2020-Results/default.aspx>> accessed 28 July 2022

⁴⁴ Miranda Sissons and Alex Warofka, ‘An Update on Facebook’s Human Rights Work in Asia and Around the World’ (*Meta*, 12 May 2020) <<https://about.fb.com/news/2020/05/human-rights-work-in-asia/>> accessed 28 July 2022

⁴⁵ Article One Advisors, ‘Assessing the Human Rights Impact of the Facebook Platform in Sri Lanka’ (2018) <https://www.articleoneadvisors.com/s/Sri-Lanka-HRIA_-_Executive-Summary_FINAL.pdf> accessed 28 July 2022

⁴⁶ Ryan Broderick and Jules Darmanin, ‘The “Yellow Vest” Riots in France Are What Happens When Facebook Gets Involved With Local News’ (*BuzzFeed*, 06 December 2018) <<https://www.buzzfeednews.com/article/ryanhatesthis/france-paris-yellow-jackets-facebook>> accessed 28 July 2022

⁴⁷ M D Conover and others, ‘Political Polarization on Twitter’ (2011) Proceedings of the Fifth International Association for the Advancement of Artificial Intelligence Conference on Weblogs and Social Media 89

with than otherwise,⁴⁸ the operators of the platforms themselves have determined that divisive content tends to attract user attention and increase time spent online.⁴⁹

The use of social media for news consumption and production, particularly with regard to political organizing and even protest activity, points to the broad range of individuals and institutions interested in that use. News consumers supplement their traditional media diet with news over social media and use the technology to support further discussion of news events and association with other like-minded people. News producers want to take advantage of the same features of social media that attract consumers, but social media also creates new opportunities for intense competition from non-traditional news media sources. The participatory capacity of social media has featured in its use in political organizing, and accordingly the medium is a matter of concern (whether positive or negative) for governments. And finally the social media companies are interested in how their services are used and the impact of those uses on the societies where the companies operate. For each of these stakeholders, the authenticity and reliability of political news on social media takes on paramount importance, especially in the context of elections.

The Fake News Problem

Struggles with the evaluation of media content as authentic or not stretches back long before the advent of social media. The idea of “unintended influence” as a kind of social harm took root as printed material of various kinds became more available and literacy spread: “The typical forms of harm that were envisaged in late medieval and early modern times could be classified as religious error, civil disaffection, or personal moral corruption.”⁵⁰ McQuail suggests that the elaboration of early admonitions against corrupting literary influences between the reigns of Henry IV and Henry VIII (and similarly in the Ottoman empire) reflected these rulers’ implicit rationale that the production scale and distribution potential afforded by printing technology increased the risk of these harms.⁵¹ Laws in the sixteenth century chased after consumer demand for “heretical and subversive literature,” foreign texts, and dissident religious publications. The diffusion of “scriptural inaccuracies” was big business, and popular works included “Hermetic writings, astrologies and books of

⁴⁸ Kevin Munger, ‘Social Media, Political Science, and Democracy’ (2019) 81(3) *Journal of Politics* 62 (reviewing some of the literature on certain negative effects of social media on the polity)

⁴⁹ Jeff Horwitz and Deepa Seetharaman, ‘Facebook Executives Shut Down Efforts to Make the Site Less Divisive’ (*Wall Street Journal*, 26 May 2020) <<https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499>> accessed 28 July 2022

⁵⁰ McQuail (n 5) 140

⁵¹ *ibid* 141

necromancy . . . reproduced in large numbers,” along with “collections of prophecies, hieroglyphics and magic practices.”⁵²

Even more obvious historical parallels to social media can be found in the early modern era, especially as news communication took on a prominent social role. In correspondence with an aspiring newspaper publisher, America’s third president Thomas Jefferson lamented the state of the early-19th Century press, the difficulty of squaring misinformation against facts within readers’ actual knowledge, and the temptation of editors to feed “the public mind habitually on slander” and salacious detail rather than mere facts and “mature consideration of all circumstances.”⁵³ Lippmann echoed strikingly similar sentiments over a hundred years later in his assessment of the press post-industrialization, describing also the persistent difficulties of distance and accountability in the formation of opinions about public affairs.⁵⁴ Radio and television presented new possibilities in news reporting beyond the capabilities of print formats. But perhaps the 20th Century’s most archetypal instance of “fake news” was the 1938 radio broadcast of the Mercury Theatre adaptation of H.G. Wells’s 1898 novel, *The War of the Worlds*.⁵⁵ Orson Welles and his company of actors used the pacing, vocal characterization, and audience expectations of actual radio news bulletins to deliver a real-time depiction of a Martian invasion of New Jersey. The performance was so convincing that terrified listeners called the police, and newspapers the next morning reported on mass panic caused by the broadcast.

The consequences of the propagation of misleading or inauthentic information have seemingly only intensified with the advance of communications technology. In 1964, U.S. President Lyndon Johnson took to the television airwaves to announce to the nation that the North Vietnamese military had attacked U.S. ships in the Gulf of Tonkin, precipitating full-scale war by the U.S. in Vietnam.⁵⁶ Within days, Johnson obtained near unanimous support for the war from the Congress via its Gulf of Tonkin Resolution. Unlike the *War of the Worlds* broadcast, for which the press held Welles to account the next morning, Johnson’s announcement of the Gulf of Tonkin incident resisted deeper scrutiny for over 40 years until related documents and transcripts were declassified. Historians, rather than journalists, were finally able to reveal the falsity of the President’s account of what had happened and

⁵² Burke (n 6) 116

⁵³ Thomas Jefferson, ‘Letter to John Norvell’ (11 June 1807) Founders Online, National Archives <<https://founders.archives.gov/documents/Jefferson/99-01-02-5737>> accessed 28 July 2022

⁵⁴ Walter Lippmann, *Public Opinion* (Simon & Schuster 1997) 18 & 29

⁵⁵ A Brad Schwartz, ‘The Infamous “War of the Worlds” Radio Broadcast Was a Magnificent Fluke’ (*Smithsonian Magazine*, 06 May 2015) <<https://www.smithsonianmag.com/history/infamous-war-worlds-radio-broadcast-was-magnificent-fluke-180955180/>> accessed 28 July 2022

⁵⁶ Pat Paterson, ‘The Truth About Tonkin’ (February 2008) 22(1) *Naval History Magazine* <<https://www.usni.org/magazines/naval-history-magazine/2008/february/truth-about-tonkin>> accessed 28 July 2022

the apparent deception of Congress by the Secretary of Defense. In the cable television era, the propensity of news media to influence foreign policy decisions in real time came to be called the ‘CNN effect’—with the caveat by Robinson that the direction of influence between the media and government goes both ways.⁵⁷ Bernal contends that the ways social media platforms like Facebook are used for news production and consumption, and the ways those platforms are structured as businesses, may mean that fake news will remain prevalent in the Internet era as well.⁵⁸

Concern about the influence of social media platforms over political division and news content particularly revolves around the elections context in western democracies.⁵⁹ Despite this concern, however, political party affiliation in the U.S. has remained relatively stable over the past decade,⁶⁰ as have news audience polarization measurements in the U.S., U.K., and elsewhere.⁶¹ A recent U.S. survey of *which* news outlets are considered trustworthy sources for political and election news notably indicated commonality between Republican- and Democratic-leaning respondents with regard to public service-oriented outlets like PBS and the BBC, while partisans tended to view the other side’s most trusted source as the least trustworthy—Fox News and CNN, respectively.⁶² Asked by a journalist what was most likely to blow a government off course, Prime Minister Harold Macmillan is purported to have replied “Events, dear boy, events”—a phrase which Macmillan may never have uttered at all, despite wide circulation of the account.⁶³ The supposed incident and utterance nevertheless captures the central role of narratives about news events, their political currency irrespective of their authenticity, and the concern authorities have over their influence.

Getting a clear picture of today’s fake news problem is complicated by the participation in its definition by interested parties and governments. In a deep sense, the ‘fakeness’ of fake news is immaterial, as what makes any news event ‘news’ is “its

⁵⁷ Piers Robinson, ‘The CNN effect: can the news media drive foreign policy?’ (April 1999) 25(2) *Review of International Studies* 301, 301-02; for a contemporary look at how political foreign policy positions manifest partisan differences in opinion polling, see Mohamed Younis, ‘China, Russia Images in U.S. Hit Historic Lows’ (*Gallup*, 01 March 2021) <<https://news.gallup.com/poll/331082/china-russia-images-hit-historic-lows.aspx>> accessed 28 July 2022

⁵⁸ Paul Bernal, ‘Fakebook: why Facebook makes the fake news problem inevitable’ (2018) 69(4) *Northern Ireland Legal Quarterly* 513, 514

⁵⁹ e.g., Hunt Allcott and others, ‘The Welfare Effects of Social Media’ (2019) National Bureau of Economic Research Working Paper 25514 <<http://www.nber.org/papers/w25514>> accessed 28 July 2022

⁶⁰ Gallup, ‘Party Affiliation’ (2022) <<https://news.gallup.com/poll/15370/party-affiliation.aspx>> accessed 28 July 2022

⁶¹ Hazard Owen (n 17)

⁶² Pew Research Center, ‘Large differences in news sources trusted by Democrats and Republicans’ (17 January 2020) <https://www.pewresearch.org/journalism/2020/01/24/democrats-report-much-higher-levels-of-trust-in-a-number-of-news-sources-than-republicans/pj_2020-01-24_media-polarization_1-01/> accessed 28 July 2022

⁶³ Elizabeth Knowles, *What They Didn’t Say: A Book of Misquotations* (OUP 2006)

dissemination, not its objective reality.”⁶⁴ Thus, as indicated by Bernal,⁶⁵ understanding the levels of organization and systematization at play in the spread of fake news can reveal much about why the political character of online activity takes on such a central role in its discussion. Press relations firms now describe the traditional media landscape in apocalyptic terms, urging the corporate sector to “take control of its own story, with every company becoming a media company.”⁶⁶ CEOs and presidents of multinationals like Microsoft describe pressure from within their own firms to take public positions on political issues.⁶⁷ As more corporations behave more like the news media, the public increasingly views business as similarly untrustworthy.⁶⁸ Businesses might instead look to the world’s “most trusted” governments (according to the PR firms) that have bucked recent trends and maintained high levels of confidence, such as Saudi Arabia, China, and the UAE.⁶⁹ In the U.S., the 2016 presidential election contest supposedly “shook American democracy to its foundations,” and the concomitant rise of online fake news has brought the nation to “a uniquely dangerous moment in history”⁷⁰—developments apparently eclipsing both the American Civil War and the Cuban Missile Crisis. Nyhan has stressed that “exaggerated fears about the influence of online information are reminiscent of past panics about the influence of television and radio” and that “we should be cautious before empowering private companies like Facebook or governments to engage in unprecedented interventions into national political debate on social media platforms.”⁷¹

Facebook came under heightened government scrutiny in early 2018, as details came to light about a data breach involving third-party companies and millions of Facebook

⁶⁴ Jacques Ellul, *Propaganda: The Formation of Men’s Attitudes* (Konrad Kellen and Jean Lerner trs, Vintage 1973) 47-48

⁶⁵ Bernal (n 58)

⁶⁶ Richard Edelman, ‘Fake News; the Neutron Bomb Explodes’ (09 November 2017) <<https://www.edelman.com/insights/fake-news-neutron-bomb-explodes>> accessed 28 July 2022 (note: a plurality of respondents indicated ‘sloppy or biased reporting’ as their definition of fake news)

⁶⁷ Arun Sudhaman, ‘Davos: Edelman Wants PR to Shift From Advocacy to Education in Fake News Era’ (*Provoke Media*, 23 Jan 2018) <<https://www.provokemedia.com/latest/article/davos-edelman-wants-pr-to-shift-from-advocacy-to-education-in-fake-news-era>> accessed 28 July 2022 (discussion with Microsoft president Brad Smith)

⁶⁸ See Edelman, ‘2021 Edelman Trust Barometer UK Findings’ (*Edelman*, 22 February 2021) <<https://www.edelman.co.uk/edelman-trust-barometer-2021-uk-findings>> (noting trust in U.K.-based companies has fallen to its lowest in 8 years)

⁶⁹ Ismaeel Naar, ‘Saudi Arabia’s government tops Edelman’s 2021 Trust Barometer report’ (*Al Arabiya*, 27 March 2021) <<https://english.alarabiya.net/News/gulf/2021/03/27/Saudi-Vision-2030-Saudi-Arabia-s-government-tops-Edelman-s-2021-Trust-Barometer-report>> accessed 28 July 2022

⁷⁰ Anthony J Gaughan, ‘Illiberal Democracy: The Toxic Mix of Fake News, Hyperpolarization, and Partisan Election Administration’ (2017) 12 Duke J Const L & Pub Pol’y 59, 60-62

⁷¹ Brendan Nyhan, ‘We Can’t Combat Fake News If We Don’t Really Understand It’ (*GEN*, 07 March 2019) <<https://gen.medium.com/we-cant-combat-fake-news-if-we-don-t-really-understand-it-caed33dbea09>> accessed 28 July 2022

users.⁷² A Cambridge University researcher had developed a ‘personality quiz’ app on the Facebook platform in 2013⁷³ and shared the Facebook user-data collected by his app with a political consultancy firm, Cambridge Analytica, in 2014.⁷⁴ That unauthorized data sharing violated Facebook’s platform policy: the quiz app was banned in 2015 after journalists exposed the scheme, and the parties were directed to delete the improperly acquired data.⁷⁵ But by 2018, inquiries by the U.K. Parliament and the Information Commissioner’s Office (“ICO”) concerning the use of personal data in political campaigning had revived interest in Cambridge Analytica’s activities, particularly in light of the ‘Brexit’ referendum campaign in the U.K.⁷⁶ Investigations by *The New York Times* and *The Observer* gathered testimony from several former employees, along with other documentary evidence, indicating that Cambridge Analytica had not complied with Facebook’s directive to delete the private data.⁷⁷ The ICO⁷⁸ and the U.S. Federal Trade Commission (“FTC”)⁷⁹ levied penalties against Facebook for its failures to protect user privacy, while Cambridge Analytica declared insolvency and ceased trading in May 2018.⁸⁰

The nature of Cambridge Analytica’s business underlies the intense interest by regulators and journalists into the data breach: Cambridge Analytica used the detailed personal information it obtained to create targeted political advertisements.⁸¹ Although the Information Commissioner found no evidence of Cambridge Analytica’s involvement in the referendum campaign,⁸² Cambridge Analytica had provided services for the 2016 U.S. presidential campaigns first of Texas Senator Ted Cruz and subsequently of Donald Trump,

⁷² Carole Cadwalladr and Emma Graham-Harrison, ‘Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach’ (*The Observer*, 17 March 2018) <<https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>> accessed 28 August 2022

⁷³ US Senate Committee on the Judiciary and US Senate Committee on Commerce, Science and Transportation, *Testimony of Mark Zuckerberg Chairman and Chief Executive Officer, Facebook* (10 April 2018) (hereinafter ‘Zuckerberg Senate Testimony’)

⁷⁴ Cadwalladr and Graham-Harrison (n 72)

⁷⁵ Zuckerberg Senate Testimony (n 73) 2

⁷⁶ Cadwalladr and Graham-Harrison (n 72)

⁷⁷ Matthew Rosenberg, Nicholas Confessore, and Carole Cadwalladr, ‘How Trump Consultants Exploited the Facebook Data of Millions’ (*The New York Times*, 17 March 2018) <<https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html>> accessed 28 August 2022

⁷⁸ Elizabeth Denham, ‘Letter to Julian Knight MP, Chair, Digital, Culture and Media and Sport Select Committee’ (Information Commissioner’s Office, 02 October 2020) <https://ico.org.uk/media/action-weve-taken/2618383/20201002_ico-o-ed-l-rtl-0181_to-julian-knight-mp.pdf> accessed 28 August 2022

⁷⁹ Federal Trade Commission, ‘FTC Imposes \$5 Billion Penalty and Sweeping New Privacy Restrictions on Facebook’ (24 July 2019) <<https://www.ftc.gov/news-events/news/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions-facebook>> accessed 28 August 2022

⁸⁰ Olivia Solon and Oliver Laughland, ‘Cambridge Analytica closing after Facebook data harvesting scandal’ (*The Guardian*, 02 May 2018) <<https://www.theguardian.com/uk-news/2018/may/02/cambridge-analytica-closing-down-after-facebook-row-reports-say>> accessed 28 August 2022

⁸¹ Cadwalladr and Graham-Harrison (n 72)

⁸² Denham (n 78)

including “designing target audiences for digital ads and fund-raising appeals, modelling voter turnout, buying \$5 million in television ads and determining where Mr. Trump should travel to best drum up support.”⁸³ Although Cambridge Analytica disputed the relevance of the misappropriated Facebook data to those efforts,⁸⁴ both the ICO and FTC have indicated that Cambridge Analytica had aggregated Facebook profile data with U.S. voter records to make predictions about individuals’ politics and personalise campaign advertisements.⁸⁵ Traditional political marketing campaigns require expensive national surveys, consumer purchase histories, and other data to build comparable voter profiles,⁸⁶ so the Cambridge Analytica incident showed that social media platforms like Facebook had developed a significantly large capacity to reach members of the public, gather information from them, and potentially influence them.

Amidst the 2020 U.S. election, an early venture capital investor in Facebook and advisor to its founder Mark Zuckerberg accused Facebook and other social media platforms of profiting from radicalization and extremism and demanded the engineers behind the technology “be held accountable for harms.”⁸⁷ But such calls for accountability may be difficult to square with the actions of various governments with long histories⁸⁸ of political media manipulation of the media, or those of committed partisans who have decided to “fight fire with fire” in the “information wars” through advocacy-journalism.⁸⁹ Furthermore, politicians around the world now use the mere notion of ‘fake news’ as a discursive device, not only to appeal to the general public but also to engage news media considerations of newsworthiness and garner press attention.⁹⁰ The U.K., which according to the U.K. Government is “recognised as a world leader in impartial and independent media,” has spent millions of pounds combatting disinformation and promoting “social cohesion” through media development in the Baltic States and Ukraine since 2016.⁹¹ Yet at home, U.K. trust in

⁸³ Rosenberg, Confessore, and Cadwalladr (n 77)

⁸⁴ Cadwalladr and Graham-Harrison (n 72)

⁸⁵ Denham (n 78); Federal Trade Commission, ‘FTC Sues Cambridge Analytica, Settles with Former CEO and App Developer’ (24 July 2019) <<https://www.ftc.gov/news-events/news/press-releases/2019/07/ftc-sues-cambridge-analytica-settles-former-ceo-app-developer>> accessed 28 August 2022

⁸⁶ Rosenberg, Confessore, and Cadwalladr (n 77)

⁸⁷ Roger McNamee, ‘Platforms Must Pay for Their Role in the Insurrection’ (*Wired*, 07 January 2021) <<https://www.wired.com/story/opinion-platforms-must-pay-for-their-role-in-the-insurrection/>> accessed 28 July 2022

⁸⁸ e.g., Carl Bernstein, ‘The CIA and the Media’ (*Rolling Stone*, 20 October 1977) <<https://www.carlbernstein.com/the-cia-and-the-media-rolling-stone-10-20-1977>> accessed 28 July 2022

⁸⁹ Theodore Schleifer, ‘Inside the new \$65 million push from progressives to compete with conservative media’ (*Vox*, 23 February 2021) <<https://www.vox.com/platform/amp/recode/22296522/acronym-democrats-courier-progressive-news-tara-mcgowan-project-for-good-information>> accessed 28 July 2022

⁹⁰ Kate Farhall and others, ‘Political Elites’ Use of Fake News Discourse Across Communications Platforms’ (2019) 13 *Int’l J of Comms* 4353, 4360

⁹¹ Foreign, Commonwealth & Development Office, ‘Counter Disinformation and Media Development Programme’ (2021)

news media of all sorts has plummeted and most news organizations are viewed as more ideological than informative.⁹² Can responsibility for developments like these fairly be laid at the feet of Zuckerberg and his employees?

The 2019 U.K. General Election

In the 2019 U.K. General Election, social media featured centrally in multiple controversies over fake news, campaign advertising, and accusations against politicians. During a televised debate in November 2019 between party leaders Boris Johnson and Jeremy Corbyn, the Conservative campaign headquarters press office changed its Twitter-verified profile and background logo to appear as 'factcheckUK' rather than its typical 'CCHQPress' brand—and then proceeded to live-tweet about the debate, including appraisals of the candidates' statements as 'facts' or not.⁹³ Online fact-checkers Full Fact, an independent charity, pointed out this campaign tactic to Twitter, and the platform warned the Conservatives that further attempts to mislead the public would result in corrective action.⁹⁴ Complaints to the Electoral Commission, which does not regulate election campaign content, resulted in a renewed call for campaigners to act responsibly.⁹⁵ Misleading campaign posts and advertisements on Facebook came under close scrutiny at the beginning of December 2019, in the run-up to polling day.⁹⁶ Facebook had a policy exempting political adverts from its third-party fact-checking programme, but Full Fact had nevertheless undertaken regular scrutiny of claims by the Conservative and Labour campaigns.⁹⁷ First Draft, an American non-profit coalition, compared Full Fact's determinations with data First Draft obtained from Facebook.⁹⁸ First Draft found that of the Conservatives' nearly 7,000 thousand ads run from December 1st to December 4th, nearly 6,000 featured claims labelled misleading by Full Fact (such as how many hospitals would be built or nurses hired by the Conservatives); while ads run by the Liberal Democrats featured far fewer misleading claims, and those run by Labour fewer still.⁹⁹ Nick Clegg, the former deputy prime minister and now an executive at Facebook, defended the company's policy to take a more permissive stance on campaign content compared to other user-generated content, citing the importance of

<[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/842428/Counter Disinformation and Media Development Programme.odt](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/842428/Counter_Disinformation_and_Media_Development_Programme.odt)>

⁹² Edelman (n 68)

⁹³ Frances Perraudin, 'Twitter accuses Tories of misleading public with "factcheck" foray' (*The Guardian*, 20 November 2019) <<https://www.theguardian.com/politics/2019/nov/20/twitter-accuses-tories-of-misleading-public-in-factcheck-row>> accessed 29 October 2022

⁹⁴ Perraudin (n 93)

⁹⁵ *ibid*

⁹⁶ Alastair Reid and Carlotta Dotto, 'Thousands of misleading Conservative ads side-step scrutiny thanks to Facebook policy.' (*First Draft*, 6 December 2019) <<https://firstdraftnews.org/latest/thousands-of-misleading-conservative-ads-side-step-scrutiny-thanks-to-facebook-policy/>> accessed 28 July 2022

⁹⁷ *ibid*

⁹⁸ *ibid*

⁹⁹ *ibid*

allowing society as a whole to make its democratic decisions, rather than the social media companies.¹⁰⁰

Yet perhaps the most striking examples of dubious social media news use in the run-up to the general election involved not only the social media companies themselves, but also the traditional media and civil society organizations. In May 2019, the Equality and Human Rights Commission (“EHRC”) announced an investigation into allegations of anti-Semitism in the Labour Party.¹⁰¹ Human Rights Watch included the Labour situation alongside the vandalism of a Jewish cemetery in France and a spike in assaults against Jewish people in Germany at the time.¹⁰² In its report on the investigation, the EHRC noted that social media was “the source of most complaints of anti-Semitic behaviour” that it investigated.¹⁰³ The Home Affairs Committee and the Community Security Trust likewise found a “sharp increase” in the number of social media-related anti-Semitic incidents, particularly during 2015-2016 and the years following—far more than had been recorded annually between 2000 and 2014.¹⁰⁴ One MP who left the Labour Party in February 2019 wrote that “[until] Jeremy Corbyn became its leader in September 2015, Labour did not have a problem with anti-Jewish racism. Today, it is institutionally antisemitic.”¹⁰⁵ In July 2019, the BBC dedicated an episode of its long-running *Panorama* series to the question of whether Labour as a party is anti-Semitic,¹⁰⁶ and in late November 2019 within weeks of the general election, BBC interviewer Andrew Neil repeatedly pressed Corbyn to apologize to the U.K. Jewish community.¹⁰⁷

Despite having swelled the ranks of Labour Party membership in 2015 and winning dozens of seats from the Conservatives in the 2017 election, Corbyn’s Labour lost the 2019

¹⁰⁰ Andrew Griffin, ‘Facebook PR boss Nick Clegg defends its decision to allow false political ads’ (*Independent*, 20 January 2020) <<https://www.independent.co.uk/life-style/gadgets-and-tech/facebook-nick-clegg-political-ads-2020-election-a9292891.html>> accessed 28 July 2022

¹⁰¹ Equality and Human Rights Commission, ‘Investigation into antisemitism in the Labour Party’ (October 2020) 17 (“EHCR”)

¹⁰² Eva Cossé, ‘The Alarming Rise of Anti-Semitism in Europe’ (*Human Rights Watch*, 04 June 2019) <<https://www.hrw.org/news/2019/06/04/alarming-rise-anti-semitism-europe>> accessed 29 October 2022

¹⁰³ EHRC (n 101) 11-12

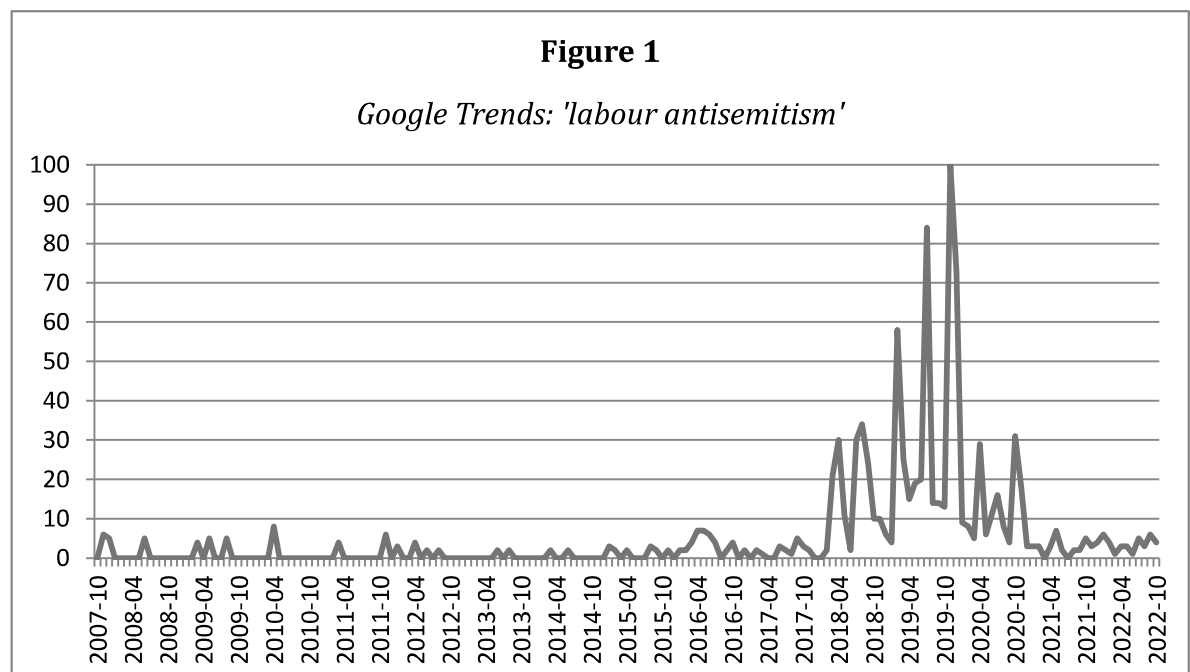
¹⁰⁴ Home Affairs Committee, ‘Antisemitism in the UK’ HC 136 (13 October 2016) 16; see also EHRC (n 101) 17 fn 2

¹⁰⁵ Joan Ryan, ‘Labour poses an existential threat to British Jews’ (*The Jewish Chronicle*, 20 February 2019) <<https://www.thejc.com/comment/comment/i-cannot-remain-in-a-party-that-poses-an-existential-threat-to-british-jews-1.480314>> accessed 29 October 2022

¹⁰⁶ BBC *Panorama*, ‘Is Labour Anti-Semitic?’ (10 July 2019) <<https://www.bbc.co.uk/programmes/m0006p8c>> accessed 29 October 2022 (the episode is not currently available for replay)

¹⁰⁷ BBC, ‘General election 2019: No apology from Jeremy Corbyn over Labour anti-Semitism claims’ (*bbc.co.uk*, 26 November 2019) <<https://www.bbc.co.uk/news/election-2019-50564965>> accessed 29 October 2022 (the episode is not currently available for replay)

general election to Johnson's Conservatives. Google Trends¹⁰⁸ shows a dramatic uptick in search interest for "labour antisemitism" beginning in 2018 and peaking in November 2019 before falling back to typical pre-2017 levels from 2021 to the present, as shown in **Figure 1**. Graeber contends that the anti-Semitism accusations weakened Labour "immensely" and were characteristically political when compared with the scant references in the British press to Conservative anti-Semitism, particularly given Johnson's own description of "'Jewish oligarchs' as controlling the global media" in comparison to Corbyn's "long history of supporting Jewish causes and [having worked] especially closely with the Haredim community."¹⁰⁹



But were the accusations 'fake news'? In April 2020, a leaked dossier prepared by Labour Party officials and a cache of emails and WhatsApp messages revealed intense factional opposition *within the Labour Party* against Corbyn.¹¹⁰ Members of the Party's senior management team privately expressed disdain and even violent sentiments towards

¹⁰⁸ Google Trends, 'Interest over time: labour antisemitism, 01 October 2007 – 01 November 2022' (Google, 2022) <<https://trends.google.com/trends/explore?date=2007-10-01%202022-11-01&geo=GB&q=labour%20antisemitism>> accessed 01 November 2022; unfortunately, Twitter does not feature a comparable 'trend' history, so I have included Google search interest as a proxy for social media traffic about the issue

¹⁰⁹ David Graeber, 'The Center Blows Itself Up: Care and Spite in the 'Brexit Election' (*New York Review of Books*, 13 January 2020) <<https://www.nybooks.com/daily/2020/01/13/the-center-blows-itself-up-care-and-spite-in-the-brexit-election/>> accessed 28 July 2022

¹¹⁰ Jon Stone, 'Anti-Corbyn Labour officials worked to lose general election to oust leader, leaked dossier finds' (*Independent*, 13 April 2020) <<https://www.independent.co.uk/news/uk/politics/labour-leak-report-corbyn-election-whatsapp-antisemitism-tories-yougov-poll-a9462456.html>> accessed 28 July 2022

Corbyn and MPs who supported him.¹¹¹ The leaked dossier says Labour Party “staffers trawled social media to find reasons to exclude [pro-Corbyn] voters from the [leadership] contest,”¹¹² This reportedly included looking for members’ posts of anti-Semitic or pro-Palestinian material.¹¹³ As noted by the EHRC when it released its findings on Labour anti-Semitism in October 2020, social media posts were central to 59 of the 70 sample complaints it investigated.¹¹⁴ The EHRC report also notes that up until 2018, the Labour Party policy on members’ social media activity meant that “even repeated sharing of antisemitic material could have escaped investigation,” blaming Labour Party leadership for failing to act.¹¹⁵ The Labour Party even settled in defamation proceedings with the BBC reporter and the *Panorama* whistleblowers, paying out a six-figure sum and issuing a formal apology for calling their accounts into question.¹¹⁶

However, subsequent investigations by the Labour Party itself¹¹⁷ and by independent journalists¹¹⁸ show that after a change in leadership (to a Corbyn ally) in 2018, the National Executive Committee of the Labour Party dramatically increased its disciplinary investigations of anti-Semitism allegations. Labour’s anti-Semitism case log indicated that during the period of Corbyn’s leadership, twenty-three percent of all complaints involved one individual complainant, while twelve percent came from an organization, Labour Against Antisemitism—with social media posts figuring centrally in many accusations against Party members.¹¹⁹ Furthermore, some of the key claims of Labour antisemitism presented in the *Panorama* story were later shown to have been handled rather uncritically by the BBC: eyewitness testimonies, leaked emails, and audio recordings of meetings flatly contradicted statements made by the *Panorama* interviewees.¹²⁰ Yet when Corbyn responded to the announcement of EHRC report’s publication in October 2020 by stating that “the scale of the

¹¹¹ Aaron Bastani, “‘It’s going to be a long night’ – How Members of Labour’s Senior Management Team Campaigned to Lose’ (*Novara Media*, 12 April 2020) <<https://novaramedia.com/2020/04/12/its-going-to-be-a-long-night-how-members-of-labours-senior-management-campaigned-to-lose/>> accessed 01 November 2022

¹¹² Stone (n 110)

¹¹³ Al Jazeera Investigations, ‘The Labour Files’ (*Al Jazeera Investigations*, 23 September 2022) <<https://www.ajjunit.com/investigation/the-labour-files/>> accessed 01 November 2022

¹¹⁴ EHRC (n 101) 85

¹¹⁵ *ibid* 85-86

¹¹⁶ Lisa O’Carroll & Jessica Elgot, ‘Labour pays out six-figure sum and apologises in antisemitism row’ (*The Guardian*, 22 July 2020) <<https://www.theguardian.com/politics/2020/jul/22/labour-pays-out-six-figure-sum-and-apologises-in-antisemitism-row>> accessed 01 November 2022

¹¹⁷ Forde Inquiry Panel, *The Forde Report* (19 July 2022) <<https://www.fordeinquiry.org/forde-inquiry-report/>> accessed 01 November 2022

¹¹⁸ e.g., Charlotte England, ‘Did BBC Whistleblowers Mislead the Public on Labour Antisemitism and Blame Their Own Failures on Corbyn?’ (*Novara Media*, 19 May 2020) <<https://novaramedia.com/2020/05/19/did-bbc-whistleblowers-mislead-the-public-on-labour-antisemitism-and-blame-their-own-failures-on-corbyn/>> accessed 01 November 2022; see also Al Jazeera Investigations (n 113) (44 cases handled versus 379)

¹¹⁹ Al Jazeera Investigations (n 113)

¹²⁰ *ibid*

problem was also dramatically overstated for political reasons by our opponents inside and outside the party, as well as by much of the media,” Corbyn was suspended from the Labour Party.¹²¹

In the wake of the April 2020 dossier leak, Labour’s National Executive Committee called for an Inquiry Panel to investigate the report and make recommendations to the Party about re-establishing itself and moving forward. The Chair of the Panel wrote the following, which encapsulates the role social media played in Labour’s purported anti-Semitism crisis:

“I understand that politics arouses passions but amplified by the echo of social media, respectful debate was replaced by strident, often coarse, tribal, and binary views. . . . People gravitated towards those with similar views and became polarized, with ill-founded speculation translated into fact and then social media propaganda. Emboldened by the possibilities and protections of social media, there was often a rush to judgement and routinely an oversimplification of the issues. This led to debilitating inertia, factionalism and infighting which then distracted from what all profess to be a common cause – electoral success.”¹²²

Forde’s statement captures key aspects of the fake news problem: how social media factors into news production and consumption, the persistent allure of sensationalism and validation, and the intensity of group identity amidst an election. As noted by the Home Affairs Committee, the majority of anti-Semitic abuse and crime “has historically been, and continues to be, committed by individuals associated with (or motivated by) far-right wing parties and political activity.”¹²³ Within the Labour Party, dozens of *Jewish* members of the Party have faced disciplinary proceedings related to allegations of anti-Semitism, most especially members of Jewish Voice for Labour.¹²⁴ Irrespective of its factual basis, the Labour anti-Semitism story had the shape and character of ‘news’ because of its currency and prominence in political discourses. Social media was integral to those factors—from its use in formulating the accusations against members to its role in legitimating the story as a subject of press discussion: if ‘everyone’ is talking about it, then the press surely has to treat it seriously. The apparent impact of social media in the 2019 general election stands in stark

¹²¹ David Renton, ‘Corbyn’s Suspension’ (*London Review of Books*, 31 August 2021) <<https://www.lrb.co.uk/blog/2021/august/corbyn-s-suspension>> accessed 01 November 2022

¹²² *The Forde Report* (n 117) 6

¹²³ Home Affairs Committee, HC 136 (n 104) 6

¹²⁴ David Hearst, ‘Geoffrey Bindman on Labour’s antisemitism scandal: “There has been political manipulation”’ (*Middle East Eye*, 26 August 2022) <<https://www.middleeasteye.net/news/labour-antisemitism-geoffrey-bindman-political-manipulation>> accessed 01 November 2022

contrast to then-Foreign Secretary Dominic Raab's comment to BBC Breakfast that "no one gives a toss about the social media cut and thrust."¹²⁵

The 2020 U.S. Presidential Election

In its 2018 investigation into alleged "Russian disinformation" attacks on American elections, the U.S. Senate Intelligence Committee relied heavily on a report by a social media research firm called New Knowledge.¹²⁶ The Texas-based research firm had attracted investment from the billionaire co-founder of LinkedIn and former government officials who had served under President Obama.¹²⁷ Despite lingering questions about the actual scale and scope of foreign interference in the 2016 election,¹²⁸ the firm's report received bipartisan acclaim from Senators like Mark Warner,¹²⁹ who has made combating disinformation one of his political priorities.¹³⁰ Somewhat ironically, however, New Knowledge was subsequently revealed to have tested its fake news theories *first-hand* by interfering in a 2017 special election campaign in the State of Alabama.¹³¹ The firm created fake Facebook accounts to message likely voters, bought retweets on Twitter, "generated phony evidence that automated Russian accounts called bots had supported [the Republican candidate]," and encouraged Alabama Republican voters to write-in a rival candidate instead of supporting the primary choice.¹³² The Republican candidate lost by less than 22,000 votes.¹³³ While

¹²⁵ Perauddin (n 93)

¹²⁶ Craig Timberg, Tony Romm, and Aaron C Davis, 'Researcher whose firm wrote report on Russian interference used questionable online tactics during Ala. Senate race' (*Washington Post*, 19 December 2018) <<https://www.washingtonpost.com/technology/2018/12/19/researcher-affiliated-with-russian-interference-senate-report-used-questionable-online-tactics-during-alabama-senate-race/>> accessed 28 July 2022

¹²⁷ Craig Timberg and others, 'Secret campaign to use Russian-inspired tactics in 2017 Ala. Election stirs anxiety for Democrats' (*Washington Post*, 06 January 2019) <https://www.washingtonpost.com/business/technology/secret-campaign-to-use-russian-inspired-tactics-in-2017-alabama-election-stirs-anxiety-for-democrats/2019/01/06/58803f26-0400-11e9-8186-4ec26a485713_story.html> accessed 28 July 2022; the founders of New Knowledge were themselves a former advisor to the Obama White House and State Department and a 15-year veteran of the National Security Agency, per Cohen (n 129)

¹²⁸ Aaron Maté, 'New Studies Show Pundits Are Wrong About Russian Social-Media Involvement in US Politics' (*The Nation*, 28 December 2018) <<https://www.thenation.com/article/archive/russiagate-elections-interference/>> accessed 28 July 2022 (pointing out report stipulations that little Russian social media activity actually pertained to the election)

¹²⁹ Dan Cohen, 'Senate report on Russian interference was written by disinformation warriors behind Alabama 'false flag operation' (*The Grayzone*, 25 December 2018) <<https://thegrayzone.com/2018/12/25/senate-report-on-russian-interference-was-written-by-disinformation-warriors-behind-alabama-false-flag-operation/>> accessed 28 July 2022

¹³⁰ Mark Warner, 'Warner: It's past time to learn from failures, adapt our laws to the internet age' (01 August 2018) <<https://www.warner.senate.gov/public/index.cfm/2018/8/warner-it-s-past-time-to-learn-from-failures-adapt-our-laws-to-the-internet-age>> accessed 28 July 2022

¹³¹ Timberg, Romm, and Davis (n 126)

¹³² *ibid*; Timberg and others (n 127)

¹³³ Timberg and others (n 127)

Democrat donors and officials have disavowed the operation,¹³⁴ the broader incident reflects the intense investment and pre-occupation of many politicians in understanding and influencing the social media news ecosystem.

The American pre-occupation with social media, fake news, and elections came to a head in October 2020, just weeks before the Presidential election. On October 14, 2020, the New York Post published a story about Democratic Party presidential candidate Joe Biden and his son Hunter Biden, who had business dealings in Ukraine during the years his father served as Vice President.¹³⁵ During the 2020 election campaign, the elder Biden had denied speaking to his son about those business dealings and denied any related conflicts of interest or wrongdoing.¹³⁶ The New York Post story, however, showed and described emails between the younger Biden and his business associates which indicated Hunter Biden had set up a meeting in 2015 between a top executive of a Ukrainian firm (where Hunter Biden sat on the board of directors) and his father, Joe Biden.¹³⁷ Then-Vice President Joe Biden subsequently pressured the Ukrainian government to fire its Prosecutor General, who had begun investigations of that company's executives.¹³⁸ Joe Biden has since rebuffed questions about the relationship in both press interviews and on the campaign trail.¹³⁹

Within hours of the publication of the New York Post story, both Twitter and Facebook took action against the Post. Twitter locked the New York Post's Twitter account, preventing the newspaper from posting anything on the social media site, and blocked users from sharing the link to the Post's article.¹⁴⁰ Twitter said it was "limiting the article's spread due to questions about 'the origins of the materials' included in the article," citing its policy on hacked materials and a "lack of authoritative reporting on where the materials included in The Post's story originated."¹⁴¹ The policy itself notably includes a recognition that "source materials obtained through leaks can serve as the basis for important reporting by news agencies meant to hold our institutions and leaders to account" and a declaration that the social media company would "defer to [news agencies'] editorial judgement in publishing

¹³⁴ Timberg and others (n 127)

¹³⁵ Emma-Jo Morris and Gabrielle Fonrouge, 'Smoking-gun email reveals how Hunter Biden introduced Ukrainian businessman to VP dad' (*New York Post*, 14 October 2020) <<https://nypost.com/2020/10/14/email-reveals-how-hunter-biden-introduced-ukrainian-biz-man-to-dad/>> accessed 01 November 2022

¹³⁶ *ibid*

¹³⁷ *ibid*

¹³⁸ *ibid*

¹³⁹ *ibid*

¹⁴⁰ Noah Manskar, 'Twitter, Facebook censor Post over Hunter Biden expose' (*New York Post*, 14 October 2020) <<https://nypost.com/2020/10/14/facebook-twitter-block-the-post-from-posting/>> accessed 28 July 2022

¹⁴¹ *ibid*

these materials."¹⁴² Twitter gave a different rationale for blocking users from sharing links to the New York Post story, calling the link “potentially harmful” and showing Twitter users who clicked the link an alert warning them that the Post webpage may be “unsafe.”¹⁴³ Facebook did not block the story outright, but reduced its distribution and put it to independent fact-checkers for further review.¹⁴⁴ Concerns about misinformation ahead of the election were at the heart of the “unprecedented” steps taken by both social media companies.¹⁴⁵

Within days of the story’s publication and subsequent social media censorship, a group of more than fifty “former senior intelligence officials” released a letter describing the Post story as having “all the classic earmarks of a Russian information operation,” albeit without any evidence to support their conclusions,¹⁴⁶ and the New York Times published an article attacking the New York Post’s editorial process, ultimately concluding that the Times could not independently verify the controversial materials.¹⁴⁷ Yet the relationship between the Bidens and Ukrainian business interests had been the subject of open discussion in the press just a year before, and anti-corruption activists in Ukraine had raised concerns that perceptions of a conflict of interest could undermine reform efforts there.¹⁴⁸ The emails presented and discussed by the New York Post appear to bear directly on lingering questions about the relationship that other outlets like the Washington Post had raised previously.¹⁴⁹ Subsequent reporting on the younger Biden’s overseas business dealings has verified the authenticity of the e-mails through expert forensic analysis and outside sources.¹⁵⁰ Had the

¹⁴² Twitter, ‘Distribution of hacked materials policy’ (October 2020) <<https://help.twitter.com/en/rules-and-policies/hacked-materials>> accessed 01 November 2022

¹⁴³ Manskar (n 140)

¹⁴⁴ *ibid*

¹⁴⁵ Kari Paul, ‘Facebook and Twitter restrict controversial New York Post story on Joe Biden’ (*The Guardian*, 15 October 2020) <<https://www.theguardian.com/technology/2020/oct/14/facebook-twitter-new-york-post-hunter-biden>> accessed 28 July 2022

¹⁴⁶ Natasha Bertrand, ‘Hunter Biden story is disinfo, dozens of former intel officials say’ (*Politico*, 19 October 2020) <<https://www.politico.com/news/2020/10/19/hunter-biden-story-russian-disinfo-430276>> accessed 01 November 2022

¹⁴⁷ Katie Robertson, ‘New York Post Published Hunter Biden Report Amid Newsroom Doubts’ (*The New York Times*, 18 October 2020) <<https://www.nytimes.com/2020/10/18/business/media/new-york-post-hunter-biden.html>> accessed 01 November 2022

¹⁴⁸ Paul Sonne, Michael Kranish, and Matt Viser, ‘The gas tycoon and the vice president’s son: The story of Hunter Biden’s foray into Ukraine’ (*Washington Post*, 28 September 2019) <https://www.washingtonpost.com/world/national-security/the-gas-tycoon-and-the-vice-presidents-son-the-story-of-hunter-bidens-foray-in-ukraine/2019/09/28/1aadff70-dfd9-11e9-8fd3-d943b4ed57e0_story.html> accessed 01 November 2022

¹⁴⁹ Morris & Fonrouge (n 135) raise many of the same points as Sonne, Kranish, and Viser (n 148)

¹⁵⁰ Craig Timberg, Matt Viser, & Tom Hamburger, ‘Here’s how The Post analyzed Hunter Biden’s laptop’ (*Washington Post*, 30 March 2022) <<https://www.washingtonpost.com/technology/2022/03/30/hunter-biden-laptop-data-examined/>> accessed 01 November 2022; Matt Viser, Tom Hamburger, and Craig Timberg, ‘Inside Hunter Biden’s multimillion-dollar deals with a Chinese energy company’ (*Washington Post*, 30 March 2022) <<https://www.washingtonpost.com/politics/2022/03/30/hunter-biden-china-laptop/>> accessed 01 November 2022

social media companies been justified in taking the actions they did concerning this particular political news story on their services?

When the New York Post was blocked and the story suppressed, opposition politicians in the U.S. immediately asked for further explanation from the platforms of what many viewed as a politically motivated intervention.¹⁵¹ The Republican Party's National Committee filed a formal complaint with the Federal Election Commission over Twitter's actions, but the Commission found in favour of the social media company, owing to the company's valid commercial reason for disallowing content that violated its policies rather than a political motivation to provide support to the Democratic presidential candidate, Joe Biden.¹⁵² According to the Commission, Twitter's head of site integrity had "received official warnings throughout 2020 from federal law enforcement that 'malign state actors' might hack and release materials associated with political campaigns and that Hunter Biden might be a target of one such operation."¹⁵³ In a 2022 interview, Facebook CEO Mark Zuckerberg commented on the Post incident, describing a prior warning to his company from the U.S. Federal Bureau of Investigation ("FBI") to be on high alert for a propaganda "dump" similar to the supposed Russian interference in the 2016 U.S. election.¹⁵⁴

By the end of March 2021, several months after the election, Twitter's then-CEO Jack Dorsey told a Congressional hearing on misinformation and social media that his company had made "a 'total mistake' by barring users from sharing" the Post's October report, and described the over two-week long lockout of the Post's account as "literally just a process error."¹⁵⁵ The Washington Post editorial board has since questioned the news media's handling of the story, including its own, asking why confirmation of the 2020 revelations only emerged over a year and a half after surfacing.¹⁵⁶ Others have pointed to the debacle as part of the on-going decline in press credibility in the U.S.¹⁵⁷ Harsher critics contend that the suppression of the story by social media and the mainstream press amounted to "election

¹⁵¹ Manskar (n 140); Paul (n 145)

¹⁵² Shane Goldmacher and Kate Conger, 'Under G.O.P. Pressure, Tech Giants Are Empowered by Election Agency' (*New York Times*, 13 September 2021) <<https://www.nytimes.com/2021/09/13/us/politics/fec-twitter-hunter-biden.html>> accessed 01 November 2022

¹⁵³ Goldmacher and Conger (n 152)

¹⁵⁴ *The Joe Rogan Experience*, '#1863 – Mark Zuckerberg' (*Spotify*, 25 August 2022) <<https://open.spotify.com/episode/51gxrAActH18RGhKNza598?si=a61ababed4584e76>> accessed 28 August 2022

¹⁵⁵ Noah Manskar, 'Jack Dorsey says blocking Post's Hunter Biden story was "total mistake" – but won't say who made it' (*New York Post*, 25 March 2021) <<https://nypost.com/2021/03/25/dorsey-says-blocking-posts-hunter-biden-story-was-total-mistake/>> accessed 28 July 2022

¹⁵⁶ Washington Post Editorial Board, 'The Hunter Biden story is an opportunity for a reckoning' (*Washington Post*, 03 April 2022) <<https://www.washingtonpost.com/opinions/2022/04/03/hunter-biden-story-is-an-opportunity-reckoning/>> accessed 01 November 2022

¹⁵⁷ Holman W Jenkins Jr., 'Media Bias and Hunter's Laptop' (*Wall Street Journal*, 12 April 2022) <<https://www.wsj.com/articles/media-spin-hunter-biden-laptop-news-trump-free-speech-censorship-stop-the-steal-big-lie-2020-presidential-election-steele-dossier-11649792758>> accessed 01 November 2022

interference” by stymying public debate over Biden’s character.¹⁵⁸ Zuckerberg’s recent reflections on the moderation process at Facebook included the following statement, which squarely frames the problem of fake news from the perspective of the social media firms, both in terms of their obligations to stakeholders and the resources which the firms can bring to bear:

“There’s also this important set of philosophical discussions which is . . . almost everyone will agree that . . . you don’t want countries basically creating networks of bots trying to convince people of stuff, you don’t want terrorism, or you don’t want child pornography . . . you don’t want people inciting literal violence. But then the question is ok, so you build these capabilities to try to find this stuff, and it’s a combination of . . . basically humans, really expert humans, and really powerful AI systems working together. But like, sometimes they get it wrong, and . . . we end up taking down accounts that we weren’t supposed to take down, and—and that sucks, right? Because then we’re kind of getting in the way of people expressing legitimate things. And . . . no system is ever going to be perfect, so the question is, . . . do you want more false positives or false negatives?”¹⁵⁹

The problem of fake news on social media is evidently more than just a problem of news on social media being ‘fake.’ Barriers to effective news use in traditional media like distance, audience taste, and language are replicated on social media as well. The scale and reach achieved by social media platforms problematizes the currency or persuasiveness of certain ideas (or even errors and mistakes) shared online. Likewise, the vast surveillance power of social media firms—the data they collect about their users and the content they post—invites the abuse of the platforms’ power by those interested in controlling news narratives. Furthermore, the loss of trust in institutions, from newspapers and television to governments and corporations, has eroded some of the resources news consumers might typically turn to in order to critically evaluate the news narratives and influence campaigns they encounter on social media. The breadth of the problem of fake news thus necessitates a broad approach to the issue in order to adequately describe 1) what happens when social

¹⁵⁸ Miranda Devine, ‘Media helped hide the real Joe Biden by censoring Hunter stories: Devine’ (*New York Post*, 28 November 2021) <<https://nypost.com/2021/11/28/media-helped-hide-the-real-joe-biden-by-censoring-hunter-stories-devine/>> accessed 01 November 2022; see also US House Committee on the Judiciary and Permanent Select Committee on Intelligence, Joint Letter to Sec’y of State Antony Blinken (20 April 2023) <https://judiciary.house.gov/sites/evo-subsites/republicans-judiciary.house.gov/files/evo-media-document/2023-04-20-jdj-mt-to-blinken-re-public-statement-on-hunter-biden-emails_0.pdf> accessed 01 December 2023 (detailing Congressional testimony indicating the story was falsely discredited by intelligence officials to assist the Biden presidential campaign)

¹⁵⁹ *The Joe Rogan Experience* (n 154)

media users encounter political news content and 2) what information about those encounters social media firms can capture and act upon.

Fake News Studies

Social science research aims to provide reliable descriptions of fake news phenomena. Researchers have previously looked closely at the use of social media platforms during election campaigns, such as the study of Twitter and Facebook use in the 2013 German Bundestag elections. That study identified a “lack of shared methods and standards in analysing electoral races with social media,” along with the difficulty of direct analysis of internal procedures or data held by the platforms, and so tracked hashtags, posts, likes, and comments and other indicia related to political topics and news stories in order to build a useful dataset.¹⁶⁰ Subsequent social media research has tended to focus either on the fake news items themselves or on fake news producers and distribution, including approaches that focus on defining fake news, examining social media reach and target audiences, and monitoring web traffic, as discussed below. My own research focuses instead on the social psychology of social media user decision making—how users’ political affiliation relates to their interactions with social media news content.¹⁶¹ By testing this phenomenon in an experimental setting, I aim to avoid some of the pitfalls encountered in research approaches that rely on arbitrary determinations of what ‘fake news’ actually is, and to provide insights into why social media users experience the online news environment as they do.

Other efforts, particularly those with an orientation toward policy solutions, hope to identify the “shades of meaning” between different terms used to describe “the accuracy and relevance of media content”—words like disinformation and misinformation commonly used in discussions of fake news.¹⁶² Such taxonomies may focus on distinguishing factors like the subjective intent of the news producer, the assessed accuracy of the story, or how systematically a story is disseminated.¹⁶³ McCright and Dunlap, for example, espouse a sociological approach that attends to “selected characteristics” of the “messengers” of fake news, such as the messenger’s realist-versus-constructivist position regarding external truth, rhetorical style, and primary audience.¹⁶⁴ To develop a more measurable econometric in

¹⁶⁰ Lars Kaczmarek and others, ‘Social Media Monitoring of the Campaigns for the 2013 German Bundestag Elections on Facebook and Twitter’ (*Leibniz Institute for the Social Sciences*, 31 March 2014) GESIS Working Papers 31

¹⁶¹ Part III, ‘The Social Psychology of Decision Making’ and Part IV, ‘Experiment Design and Results’

¹⁶² Caroline Jack, ‘Lexicon of Lies: Terms for Problematic Information’ (*Data & Society Research Institute*, August 2017) 1 <<https://datasociety.net/library/lexicon-of-lies/>> accessed 28 July 2022

¹⁶³ *ibid* 2-5

¹⁶⁴ Aaron M McCright & Riley E Dunlap, ‘Combatting Misinformation Requires Recognizing Its Types and the Factors That Facilitate Its Spread and Resonance’ (2017) 6 *J of Applied Research in Memory and Cognition* 389, 389-90

their study, Allcott and Gentzkow narrowly defined fake news as “news articles that are intentionally and verifiably false, and could mislead readers.”¹⁶⁵ A narrow definition consequently excludes a number of “close cousins of fake news” that would be included in broader taxonomies of problematic social media content:

“1) unintentional reporting mistakes, . . . ; 2) rumors that do not originate from a particular news article; 3) conspiracy theories (these are, by definition, difficult to verify as true or false, and they are typically originated by people who believe them to be true); 4) satire that is unlikely to be misconstrued as factual; 5) false statements by politicians; and 6) reports that are slanted or misleading but not outright false”¹⁶⁶

Notably, survey respondents have indicated that biased or slanted reporting is most commonly what they mean by ‘fake news’, according to Edelman.¹⁶⁷ And “false statements by politicians” would seemingly be an exception that swallows the rule: statements by politicians, being newsworthy, become news stories; and politicians, knowing this, intentionally say false things in order to mislead news audiences.

At their roots, narrow definitions like the one used by Allcott and Gentzkow still buy their narrowness with broad assumptions—assumptions that researchers (and in turn, regulators) can divine an author’s intent, that a proposition can be readily verified, that a given reader is more or less likely to be misled by a particular story. Bernal further points out that the emphasis on defining or differentiating ‘fake news’ obscures the key features of the platforms that facilitate the spread of social media content in general: data analysis, user profiling, and targeted advertising.¹⁶⁸ Marketing scholarship in particular has highlighted the synergistic relationship between fake news stories and online advertising, which propagates without the same judicious control of ad placement exercised by brand managers using traditional media channels.¹⁶⁹ The structure of social media incentivizes making even ‘real’ news perform more like ‘fake’ news: “[R]eliance on high levels of viewership for legitimacy and financial success has resulted in the proliferation of ‘clickbait’, articles that often sacrifice objectivity for increased viewership through attention-grabbing titles designed to induce clicks.”¹⁷⁰ Accordingly, media watchdog groups have also urged careful consideration

¹⁶⁵ Hunt Allcott and Matthew Gentzkow, ‘Social Media and Fake News in the 2016 Election’ (2017) 31(2) J Econ Perspectives 211, 213

¹⁶⁶ *ibid* 214

¹⁶⁷ Edelman (n 66)

¹⁶⁸ Bernal (n 58) 515

¹⁶⁹ Adam J Mills, Christine Pitt, and Sarah Lord Ferguson, ‘The Relationship between Fake News and Advertising: Brand Management in the Era of Programmatic Advertising and Prolific Falsehood’ (March 2019) Journal of Advertising Research 3, 4-5

¹⁷⁰ Wilton Park, ‘#FakeNews: innocuous or intolerable?’ (Wilton Park Report WP1542, 15-17 February 2017)

of regulatory proposals relying on inapt definitions of ‘fake news’ or similar substitute terms and encouraged a balanced appraisal of business models and actual technological capabilities.¹⁷¹

Although Guess, Nyhan, and Reifler rely on Allcott and Gentzkow’s pre-judged news articles as a jumping-off point, the trio’s study directly tracked American participants’ web traffic during the 2016 U.S. presidential election, allowing for the comparison of participants’ use of Facebook and Twitter with visits to websites that hosted (among other content) the identified fake news articles.¹⁷² Guess, Nyhan, and Reifler inferred an association between higher Facebook usage and more fake news-coded website visits.¹⁷³ Based on URL and time data, they found that participants had accessed a Facebook URL within 30 seconds prior to over 20 percent of visits to fake news-coded websites, compared to Google, Twitter, or webmail accesses prior to fewer than 10 percent of site visits combined.¹⁷⁴ Moreover, overall exposure to such websites appeared both very limited and very selective: Articles on the fake news websites constituted only a small proportion of news articles read during the period overall, and only a small proportion of the participant population constituted most of the readers of those websites.¹⁷⁵

Adding more publishers and including manually annotated categories for more or less “flawed” editorial processes, Grinberg and others studied Twitter users’ exposure to fake news sources across an expanded (though overlapping) time period relative to the Guess, Nyhan, and Reifler study.¹⁷⁶ Like that investigation, the Grinberg study also observed that “the vast majority of fake news shares and exposures were attributable to tiny fractions of the population.”¹⁷⁷ They also observed a proportional increase in exposures to fake news sources as participants’ Twitter exposures to political news overall increased, suggesting that “a stronger selective exposure process exists for individuals with greater interest in politics”¹⁷⁸ similar to Guess, Nyhan, and Reifler’s Facebook findings.¹⁷⁹ Knowing more about how often and *which* social media users may be exposed to fake news can help policymakers come up with responses that are tailored to what the research indicates. Demographic

¹⁷¹ Wilton Park (n 170)

¹⁷² Andrew Guess, Brendan Nyhan, and Jason Reifler, ‘Selective Exposure to Misinformation: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign’ (9 January 2018) 1, 3 <<https://www.dartmouth.edu/~nyhan/fake-news-2016.pdf>> accessed 02 June 2020

¹⁷³ *ibid* 8-9

¹⁷⁴ *ibid*

¹⁷⁵ Guess, Nyhan, and Reifler (n 172) 4-5 and 10-11

¹⁷⁶ Nir Grinberg and others, ‘Fake News on Twitter during the 2016 U.S. presidential election’ (2019) 363 *Science* 374

¹⁷⁷ *ibid* 375 (discussing the outsize role of “supersharers”: and “superconsumers” in fake news source traffic); *ibid* 377

¹⁷⁸ *ibid* 375

¹⁷⁹ Guess, Nyhan, and Reifler (n 172) 5

characteristics like a Facebook user's age may provide substantial explanations of fake news sharing behaviours, which in itself suggests an array of approaches like digital literacy training and focused interventions.¹⁸⁰

Perhaps most importantly, observational studies indicate that the overall risk of fake news proliferation on social media may in fact be quite low.¹⁸¹ Studies of web traffic and Facebook engagement in Italy and France likewise have found that on average, false news outlets exhibit limited reach and draw user attention for a shorter time relative to established news brands.¹⁸² Nyhan reports that the reach of fake news in the 2018 U.S. elections appears to have fallen relative to his findings from 2016, and that online echo chambers remain "relatively rare."¹⁸³ And Nyhan also points out that Facebook's specific role in 2016 fake news-web traffic was no longer apparent in 2018—suggesting that the platform had made changes that significantly altered what news sources its users accessed.¹⁸⁴ That fact, along with the relative inscrutability of how the social media 'sausage' gets made, highlight a profound challenge in studying the fake news problem through observation: researchers must often rely on inference and supposition about the intervening role of the platforms themselves in promoting or demoting the fake news under scrutiny. How much of the 2016 fake news traffic arose 'naturally' in the course of users' online browsing, and how much should be credited to the suggestions and search results determined by the platforms?

Studies like these do tell us a lot about the fake news problem. Overall, social science research approaches provide a 'reality check' for policymakers by highlighting the hidden complexities of social problems and evaluating the results of policy interventions. Running across a dubious headline does not appear to cause brainwashing: "no credible evidence exists that exposure to fake news changed the outcome of the 2016 election" in the U.S., for example¹⁸⁵ Yet 'fake news' as a self-contained problem, whether described as the threat of "real disinformation" by foreign foes or as a rhetorical cudgel used by domestic authoritarians, has nevertheless become the centrepiece of online media restrictions passed

¹⁸⁰ Andrew Guess, Jonathan Nagler, and Joshua Tucker, 'Less than you think: Prevalence and predictors of fake news dissemination on Facebook' (2019) 5(1) *Science Advances* <<https://www.science.org/doi/10.1126/sciadv.aau4586>> accessed 28 July 2022 (*contra* younger 'digital natives', older cohorts may be 'digital immigrants' struggling to integrate)

¹⁸¹ *ibid* ("sharing fake news was quite rare during the 2016 U.S. election campaign")

¹⁸² Richard Fletcher and others, 'Measuring the reach of "fake news" and online disinformation in Europe' Reuters Institute for the Study of Journalism at the University of Oxford (February 2018)

¹⁸³ Brendan Nyhan, 'Why Fears of Fake News Are Overhyped' (*GEN*, 04 February 2019)

<<https://gen.medium.com/why-fears-of-fake-news-are-overhyped-2ed9ca0a52c9>> accessed 28 July 2022

¹⁸⁴ *ibid*

¹⁸⁵ *ibid*

or proposed in dozens of countries in the past six years.¹⁸⁶ In its Joint Declaration on Freedom of Expression and “Fake News”, Disinformation and Propaganda, the United Nations emphasized the continued importance of appropriate scrutiny for such restrictions, “namely that they be provided for by law, serve one of the legitimate interests recognised under international law, and be necessary and proportionate to protect that interest.”¹⁸⁷ As social media broadens public access to exercising rights of speech and assembly, more governments may come to distrust or fear the public’s new capabilities to discuss and dispute official accounts of events, even in “free” societies,¹⁸⁸ and Hintz has discussed some of the ways that dissident speech is vulnerable to governmental and commercial pressures on social media platforms to censor users.¹⁸⁹ U.N. Special Rapporteur David Kaye offered the following perspective to the Human Rights Council on how disinformation and propaganda affects overall public trust in media and government institutions:

“The public sees hate, abuse and disinformation in the content users generate. Governments see terrorist recruitment or discomfiting dissent and opposition. Civil society organizations see the outsourcing of public functions, like the protection of freedom of expression, to unaccountable private actors.”¹⁹⁰

In this Part, I have discussed where social media sits in a broader landscape of communications technologies, particularly through social media use for political news and organizing activities. Examples like the Cambridge Analytica scandal, the U.K. Labour anti-Semitism accusations, and the suppression of the New York Post exposé highlight the intensity of interests at stake for citizens, governments, and the social media companies alike when social media-driven narratives bear on elections and the democratic process. Lastly, I have discussed some of the research approaches which aim to enrich the context in which fake news phenomena are addressed.

The goal of my Thesis is to evaluate potential policy responses to fake political news under development in Britain and the United States. Public policy concerning social media tends to reflect specific perspectives on the role of government in communications systems,

¹⁸⁶ Eliza Mackintosh, ‘Analysis: No matter who wins the US election, the world’s “fake news” problem is here to stay’ (CNN, 25 October 2020) <<https://edition.cnn.com/2020/10/25/world/trump-fake-news-legacy-intl/index.html>> accessed 28 July 2022

¹⁸⁷ UN Special Rapporteur on Freedom of Opinion and Expression, OSCE Representative on Freedom of the Media, OAS Special Rapporteur on Freedom of Expression, and ACHPR Special Rapporteur on Freedom of Expression and Access to Information, ‘Joint Declaration on Freedom of Expression and “Fake News”, Disinformation and Propaganda’ FOM.GAL/3/17 (Vienna, 3 March 2017)

¹⁸⁸ Shirky (n 29)

¹⁸⁹ Hintz (n 41)

¹⁹⁰ United Nations General Assembly ‘Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression’ (6 April 2018) UN Doc A/HRC/38/35

ranging from preferences for the platforms to manage user-generated content themselves to preferences for more direct state intervention to curtail unwanted political discourses. Across these perspectives, the social media platforms are typically envisioned as retaining a central role in the implementation of public policy, often owing to the scale and complexity of their operations and the technical expertise the platforms hold.

Based on some of the challenges I identified in other research approaches discussed in this Part, I considered whether a broader approach to social media as a media ‘system’ might help identify research questions more closely linked to the potential efficacy of public policy initiatives concerning fake news. In the next Part, I discuss my use of media systems analysis as a framework for examining social media in terms of both its structure and its function—the elements that describe its operation and the roles we may envision for it. Specifically, I wanted a way to better answer 1) how and where ‘editorial’ decisions about fake news are made in the social media system, and 2) what may be missing or obscured from our assumptions about how these decisions can be shaped to a desired public function. By taking this approach, I aim to identify ‘sticking points’ or mismatching between how social media actually works and what policy proposals intend to achieve. I subsequently describe a key sticking point in the operation of large scale social media platforms: the automation of content moderation and its reliance on various data inputs to identify and respond to objectionable social media content like fake news.

Key Takeaways

- Use of social media for news gathering and sharing has become highly instrumental in political and civic engagement, including protest and dissident movements
- Concerns about social media news authenticity are wrapped up with concerns about influence over public opinion—and especially elections
- Studies of ‘fake news’ follow a host of conceptual and methodological approaches, producing conflicting accounts of the phenomenon and response recommendations

Part Two: Approaching Fake News as a Media System Problem

This second Part presents the analytic framework I have adopted for considering whether policy proposals addressing fake news on social media will be fit for purpose. I wanted to develop an account of social media that accurately describes how the platforms operate and evaluate various policy proposals for addressing fake news accordingly. Insights gained from examining the structural elements of the social media system (especially ones that might be subjected to policy action) can help us understand whether policy proposals are meaningfully geared towards the actual operation of social media. What public role we envision for social media in turn suggests which policy proposals may tend to be more or less effective, as the history of media policy contains a wealth of examples better or worse aligned with differently structured media systems.

In the first section of this Part, I explore looking at social media as a media system. This exploration aims to account for the characteristic structural elements of social media relevant to my Thesis, such as the content distribution mechanisms employed by the social media platforms. I do so by looking at how other media like newspapers have been interpreted as media systems and by comparing social media to highlight key similarities, differences, and interrelations with other parts of society, such as the legal system. In the second section of this Part, I take a closer look at social media platform operations, particularly with regard to massive platforms YouTube, Facebook, and Twitter. The policies and processes developed by these platforms are centrally important to understanding how social media networks respond to public obligations. I conclude this section with a review of studies on the efficacy of content moderation at scale, providing a deeper look at a unique structural element of the largest social media platforms.

The aim of this Part is to build on Part One's discussion of the problem of fake news by examining how stakeholders' assumptions, priorities, and capabilities shape the formulation and implementation of 'solutions' to fake news. The nature of the data about news content collected by the platforms, and how they operationalize it, becomes especially important where legal and regulatory regimes require or direct its use. Because of the political nature of elections news, I have surmised a potential relationship between user-generated data about news content and social media users' political group membership. In Part Three of this Thesis, I will discuss social psychological factors which may act as a substantial constraint on the efficacy of fake news moderation. Such a constraint carries its own set of implications for public policy initiatives which rely on big data-driven content moderation systems, which I discuss further in Parts Four and Five.

Social Media and Media Systems

What does social media have to do with traditional media? Isn't it part of the information technology revolution? Proprietary algorithms, hard at work behind the scenes, bring us the latest and greatest content—powered by big data and the wisdom of the crowds. Tinkering with computer code is surely the realm of computer programmers and not that of laypersons or (perish the thought) politicians. But if we were to regulate stage magicians, surely we would insist they tell us in detail how their illusions are actually accomplished. From our lay perspective, limited as we are by the social media magicians' prerogative to not divulge any trade secrets about the algorithmic magic, how can we even begin to scrutinize social media? Much like people in the 19th Century faced with the marvels of electricity, we can either choose to be mystified (if occasionally delighted) by the technology—or we can test our assumptions and reasoning about it through observation, logic, and experimentation. In this section, I draw on analyses of past media systems for grounds to describe and evaluate the structures and functions of contemporary social media.

As stated by Hardy, the concept of a 'media system' describes "features in order to go on to analyse and question how they may interconnect and relate to one another."¹⁹¹ Like the circulatory system in our bodies, the media system interacts with other organs (like the political system) that make up our social and cultural whole. Past media technology developments have been examined according to, *inter alia*, their social function—by critics, scholars, and decision makers with varying degrees of preconception about what roles media ought to fulfil. McQuail's survey of various sociologists' theories about the role of media in society reported the functions of "collecting and publishing information about social reality" and correlating communications to facilitate "cohesion, interpretation, and . . . order," to provide continuity of social order over time, and to entertain.¹⁹² In the American context, scholars have identified the pro-social role of mass communication technologies as "most critical"¹⁹³ in holding together a large society, and have promoted "civic unity and national consensus" as well as "civic consciousness" and "the creation of a sense of [national] identity" as important aims for mass communication.¹⁹⁴ Pool described how the press "serves as a major instrument of communication in the policy-making process by linking different parts of the bureaucracy, the legislative and executive branches, as well as the government and the

¹⁹¹ Jonathan Hardy, *Western Media Systems* (Routledge 2008) 6

¹⁹² Denis McQuail, *Media Accountability and Freedom of Publication* (OUP 2007) 49-50

¹⁹³ Douglass Cater, 'Toward a Public Philosophy on Government-Media Relations' in William L Rivers & Michael J Nyhan (eds), *Aspen Notebook on Government and the Media* (Praeger London 1973) 3, 11

¹⁹⁴ Ithiel de Sola Pool, 'Newsmen and Statesmen: Adversaries or Cronies?' in William L Rivers & Michael J Nyhan (eds), *Aspen Notebook on Government and the Media* (Praeger London 1973) 12, 17, 18

people” in order to “help the government carry through national policies.”¹⁹⁵ As discussed in Part One, however, social media’s function seems to be rather unsettled, with some viewing it as an emancipatory marvel and others viewing it as a disruptive scourge.¹⁹⁶

While a media systems analysis aims to describe the elements that make up something like broadcasting or social media, it does not take the system’s structures as fixed elements. Our picture of newspapers and their function has changed over time: Winston suggests that the tendency towards concentration and standardization in the mid-nineteenth century rise of mass circulation newspapers was rooted in the wider industrialized mode of that society.¹⁹⁷ The “penny press” of the 1830s formed the rudiment of British cultural production on an industrial scale which matured as newspaper ownership became more concentrated in corporate hands into the 20th Century.¹⁹⁸ Fisher noted that the functions of the media are “neither frozen nor self-determined,” stating,

“It is for society to decide what functions we would like to have the media perform, and then we must do what we can—by persuasion, by incentives, or by regulation—to have those responsibilities accepted and ably performed. . . . It may be, however, that if the media consistently fail to perform functions that society believes they should perform, legal techniques will and, perhaps, should be used to help bring about the desired performance.”¹⁹⁹

The role of the newspaper publisher as an active participant in society emerged in “more free urban centres (such as London, Hamburg, and Amsterdam), wherever business or politics or law was practiced” and in Europe especially where democratic movements concentrated during the Enlightenment.²⁰⁰ In terms of media studies, this situation of newspapers is often conceptualized as part of the ‘public sphere’ notion espoused by Habermas, a “realm of our social life in which something approaching public opinion can be formed.”²⁰¹ McQuail reiterates Habermas’ notion of the public sphere in suggesting that the modern newspaper can be considered “as itself a kind of public space or arena, where varied ideas and information are made public, compared, and debated.”²⁰² The public sphere in

¹⁹⁵ Pool (n 194) 17

¹⁹⁶ Part I, ‘Social Media News Use’

¹⁹⁷ Brian Winston, *Media technology and society: a history* (Routledge 1998) 77

¹⁹⁸ Eric Louw, *The Media and Cultural Production* (SAGE 2001) 37

¹⁹⁹ Roger Fisher, ‘The Public’s Right to Know: Functions and Malfunctions of the Media’ in William L Rivers & Michael J Nyhan (eds), *Aspen Notebook on Government and the Media* (Praeger London 1973) 78, 79

²⁰⁰ McQuail (n 192) 35; see also Louw (n 198) 44

²⁰¹ Jurgen Habermas, “The Public Sphere: An Encyclopedia Article (1964)” (1974) 3 *New German Critique* 49-55; cf. Bart Cammaerts, ‘Radical pluralism and free speech online: the case of North Belgian extreme right discourses’ (2009) 12(6) *Int’l J Cultural Studies* 555 (arguing instead for a pluralistic/agonistic conception of social media and democracy)

²⁰² McQuail (n 192) 35

turn has been carried forward in discussions of online media, but as McCullagh noted in 2003, the potential of Internet communications to be integrated into democratic processes would be contingent on public policies that “widen access to the Internet across social, economic, and organisational barriers” rather than treat online discussion as just another consumer service.²⁰³

The relationship between media systems and laws likewise can be traced over time by examining how the law has changed as we have adopted or shed various ways of using communications technologies.²⁰⁴ Solove notes that “New technologies rarely give rise to questions we have never addressed before. More often they make the old questions more complex. . . . Although modern information technology has revolutionized how we record and spread information, we have experienced similar revolutions in the past.”²⁰⁵ As shown at length by Hadfield, the examination of historical law-systems provides an informative catalogue of what has worked—what rules have been characterized by common understanding and collectively respected and enforced.²⁰⁶ The formal laws that have been developed around communications technologies were at times better or worse at achieving aims that are recognizably useful to the culture in which they emerged, and whether those laws remained valuable as circumstances changed in turn speaks to their quality as law-systems.²⁰⁷

The durability over the centuries of the notion of ‘the press’ testifies to the fundamental character of Gutenberg’s advancement: the name of the device allowing the reliable, accurate reproduction of a body of written work, quickly and at scale, became a metonym for the people who used it to record and share knowledge of the events of their time. By contrast, consider how the language we use to describe Internet communications belies the iterative nature of what we do online: we *send* an *e-mail*, *type* in *chatrooms*, *post* on *bulletin boards*, and *scroll* through *webpages*, rather than undertake some activity that would demand a neologism. ‘Tweeting’ something might come close, but the word is a branding exercise: you would never ‘tweet’ something on Facebook or YouTube, only within the walled garden of Twitter. ‘Livestreaming’ on the other hand speaks to the immediacy of electronic communications, but so too did ‘live’ broadcasting in the days of radio. So what is social media ‘like’ as a media system? Is it completely different from what’s come before?

²⁰³ Karen McCullagh, ‘E-Democracy: Potential for Political Revolution?’ (2003) 11(2) *International Journal of Law and Information Technology* 149, 160

²⁰⁴ This framework is discussed further *infra*, ‘Looking Ahead: Continuities and Discontinuities’

²⁰⁵ Daniel Solove, *The Future of Reputation: Gossip, Rumor, and Privacy on the Internet* (Yale UP 2007) 105-06

²⁰⁶ Gillian K Hadfield, *Rules for a Flat World*, ‘The Invention of Law’ (OUP 2016) 29-57

²⁰⁷ Chris Reed, ‘How to make bad law: Lessons from Cyberspace’ (2010) 73(6) *Modern Law Review* 903, 903-04

In the following pages, I look at some of the ways policymakers have responded to questions of media influence across time and technological developments. Each phase can be considered relative to what our picture of social media is like as a media system. Is social media simply like the printing press, enabling the mass production of information by many (often unaccountable) people? Is it more like the telegraphs system, providing rapid distribution across a vast centralized network? Or is it like contemporary print or broadcast media, leveraging expertise and technology to deliver ‘all the news that’s fit to print’? Social media users can generate and upload content (including news) to large platforms like Facebook and YouTube, which rely on technical systems to make user-generated content available to other users: Are users thus like journalists and audiences, and the platforms like editors—or are the platforms more like delivery services and newsagents? Rather than being just ‘like’ a particular system of media producers, distributors, and consumers, social media may actually continue some aspects that characterize certain media systems (like centralized distribution) and shed others (like liability concerns or professionalization). By looking at what structures in preceding media systems were addressed as dysfunctional, we can better identify what to look for in the social media system that might be treated the same way—and also note where responses are tied to structures that are less relevant (or scarcely developed) in social media.

Medieval Responses to Media Influence: Authorial Accountability, Licensure

While the thought of life before social media might seem impossibly dull to anyone today who can’t stop checking Twitter or Instagram, the information economy of the Middle Ages would leave them utterly shocked by its relative austerity. By the mid-Fifteenth Century, the technologies of “production, reproduction, and transmission of communication” had “scarcely progressed since Roman times,” and the concepts of “authorship, publication, and influence were scarcely more advanced.”²⁰⁸ The conditions shaping and constraining information in the medieval period included the irregularity of dialects between regions (even across relatively short distances), the difficulty of safely traveling more than a few miles, lack of literacy outside the clergy, and a general reliance on memory and oral testimonies.²⁰⁹ Any rare bit of medieval political news shared by the town crier was as likely to be misheard, misremembered, mistranslated, or misunderstood so as to be completely worthless to the average peasant or burgher, who might not have cared about it anyway unless it suggested an eminent Viking raid.²¹⁰

²⁰⁸ McQuail (n 192) 24-25

²⁰⁹ James Burke, *The Day the Universe Changed* (rev edn, Little, Brown 1995) 91-96

²¹⁰ The Medieval period was not free from consequential fake political news, of course—for example, Pope Gregory IX’s monks spread rumours of the death of Emperor Frederick II during Frederick’s crusade in

The arrival of the printing press and affordable paper would profoundly change how readily written information could be reproduced and accessed. According to Burke, “Printing was to bring about the most radical alteration ever made in Western intellectual history, and its effects were to be felt in every area of human activity.”²¹¹ The technological development of printing influenced the structure of book production and the way in which printing was used to transmit information: “Historical accounts of the impact of printing tend to emphasize the extent to which the whole book trade was organized on very commercial lines, often needing quite a lot of capital and having small-scale monopoly tendencies, responding to new opportunities in a very entrepreneurial way.”²¹² The commercial character of early printing ventures may also have arisen in satisfaction of its first customer base: the early fifteenth-century merchants who increasingly sought “rational, factual information” upon which to base business decisions.²¹³ Also suggestive of this is the fact that printing spread quickly throughout Europe in “centres of business, the sites of royal courts or the headquarters of banking organisations”²¹⁴ rather than university cities or religious sites.

Like the printing press, social media provides its users with a tool to create and replicate content. But social media has not fundamentally elevated us from a medieval level of communication characterized by information scarcity and high costs of accurate reproduction—that tectonic shift had already been accomplished long before the Internet. Nor has social media produced a necessarily unique overload of information. Already by the early 20th Century, Lippmann had identified barriers which beset the newspaper industry in its heyday that have persisted even as news production and consumption can be done online. These include the cost of primary news gathering,²¹⁵ the irreducible time required (or actually taken) to read the news,²¹⁶ the paucity of commonalities among mass audiences,²¹⁷ and the difficulty of securing a reader’s attention.²¹⁸ While social media has arguably made some of those problems worse, such challenges should not be considered novel.

Moreover, printing technology reflected a disjunction with existing social and legal technologies that focused on the protection and reproduction of the social order achieved through feudalism. Early printing workshops brought together merchants, investors,

1229, which were promptly dispelled upon the Emperor’s successful return (to the Pope’s chagrin); see Will Durant, *The Story of Civilization, Vol. IV: The Age of Faith* (Simon & Schuster 1950) 716

²¹¹ Burke (n 209) 91-96

²¹² McQuail (n 192) 29

²¹³ Burke (n 209) 108

²¹⁴ *ibid* 113

²¹⁵ Walter Lippmann, *Public Opinion* (Simon & Schuster 1997) 32

²¹⁶ *ibid* 39-40

²¹⁷ *ibid* 110-11

²¹⁸ *ibid* 222-23

translators, craftsmen, and artists in a new venture, outside the guild system and church doctrines that restricted practices in other domains.²¹⁹ In other words, it empowered new sites for the production of information, culture, and influence outside the existing power structures of the established medieval social order. The increased individual and collective participation in new discourses that early Internet theorists²²⁰ hoped would happen with the World Wide Web in the 20th and 21st Centuries actually *did* happen in the 15th and 16th Centuries as printing technology spread—yet the pace of information exchange over great distances would remain constrained until distribution technologies caught up, as discussed below.

While the technology of printing improved in quality between the sixteenth and eighteenth centuries, “the scale of production did not change very much” during that time, according to McQuail.²²¹ What did change during those years was the way members of European societies organized their way of thinking—that knowledge is accumulated and built upon what is proved and agreed—and sought to expand their awareness of people and events in different regions.²²² Printing in turn intensified communication among intellectuals, specialists and researchers as the body of shared knowledge grew and shifted.²²³

The notion of authorial accountability itself also developed alongside early printing technology rather than with Twitter: McQuail observed that “in early medieval scholarship and science, both oral and written communication were centred on the transmission of truth or authoritative accounts.”²²⁴ Feudal systems of government depended on loyalty and reputation, and so rumour and slander were long viewed by the Crown and Church as systemic threats.²²⁵ Helm and Nasu date legal sanction against fake news back to Edward I’s proclamation in 1275 concerning “any false News or Tales, whereby discord, or occasion of discord or slander may grow between the King and his People, or the Great Men of the Realm.”²²⁶ But with the advent of printing, the role of the author (especially where the author transgressed governmental interests) became “more clearly defined and recognized” by a “mixture of factors” that included legal accountability—which would be extended also to

²¹⁹ Burke (n 209) 114

²²⁰ Discussed in Part I, ‘Social Media News Use’

²²¹ McQuail (n 192) 30

²²² Burke (n 209) 118

²²³ *ibid* 123

²²⁴ McQuail (n 192) 140

²²⁵ Burke (n 209) 98

²²⁶ Rebecca K Helm and Hitoshi Nasu, ‘Regulatory Responses to “Fake News” and Freedom of Expression: Normative and Empirical Evaluation’ (2021) 21 Human Rights Law Review 302, 322 (citing *De Scandalis Magnatum* 1275 3 Edw 1 at ch 34) fn 126

publishers.²²⁷ Burke suggests more strongly the relationship between the accurate, automatic reproduction provided by printing technology and the emerging concept of authorship: “For the first time a writer could be sure of reaching a wide readership which would hold him personally responsible for what he had written.”²²⁸

In the time of the Tudors and Stuarts, information technologies like printing and theatre had expanded the influence of communication beyond word of mouth and rumour, while governmental techniques like bureaucracy had supplanted the old feudal systems. Henry VIII and his successors accordingly developed a range of legal tools to constrain the use and spread of media, whether in the audio-visual form of theatre for illiterate audiences or in print format for elite audiences. During the reign of Edward VI, “the first official act which attempted to control the stage by instituting a definite system of censorship, which required license for the printing, selling, or playing of interludes” (as stage plays were called) was instituted.²²⁹ As the sixteenth century progressed, the English monarchs centralized and formalized control of printing and distribution, placing responsibility “into the hands of the Stationers’ Company, a trade cartel, incorporated in 1557, which regulated virtually all aspects of book printing, publishing, and selling.”²³⁰ By the time of Milton’s *Areopagitica* in 1644, authors had certainly begun to feel the yoke of this responsibility as state licensing and censorship were brought to bear on them, and the hand of government control over written communication would be felt in Europe well into the 19th Century.²³¹

The endurance of licensure schemes reflects in no small way their utility to those in authority—they are an effective means of protecting centralized power from a critical or disruptive populace. Louw writes that the licensing and censorship model arose in response to the emerging power of the European middle class, the free citizens of the growing urban areas of Western Europe and their growing trade and printing-press driven communication networks.²³² These networks presented potential threats to European aristocracies and to the Church as “sites from which potentially destabilizing discourses could emerge.”²³³ Concern over these “destabilizing discourses” follows logically from the importance of reputation in the operation of the feudal systems from which the European monarchies and early nation-states were emerging.²³⁴ Opposition to such restrictions on speech and

²²⁷ McQuail (n 192) 31

²²⁸ Burke (n 209) 116

²²⁹ Harold Gardiner, *Mysteries’ End: An Investigation of the Last Days of the Medieval Religious Stage* (Yale 1946) 60

²³⁰ McQuail (n 192) 30

²³¹ Louw (n 198) 44

²³² *ibid*

²³³ *ibid*

²³⁴ *cf.* Elon Musk, ‘Status’ (01 November 2022) <<https://x.com/elonmusk/status/1587498907336118274>> accessed 30 November 2023 (describing social media platform Twitter’s blue checkmark ‘verification’)

publication would become a key feature of various 18th and 19th century revolutionary movements in Europe²³⁵ and a cornerstone of American jurisprudence on free expression.²³⁶ Fisher notes that in the U.S., the “basic guarantee of the First Amendment is against prior restraint, and for good reason. The censorship to which our forebears primarily objected was that which required a man to take his newspaper or book to a public official for approval in advance of publication.”²³⁷

For colonial pamphleteers and broadsheet printers in America, the aim of the First Amendment reflected their recent experiences as the targets of government repression, just as other elements of the U.S. Bill of Rights reflected various privations by the Crown and Parliament against the colonists in the years before the Declaration of Independence in 1776.²³⁸ Yet by 1798, the Federalist-controlled Congress passed the Sedition Act, criminalizing the publication of “any false, scandalous, and malicious writing” about the U.S. Government.²³⁹ Its language would have appeared quite sensible to the English court that, nearly a century before, had decided *Turchin* in favour of suppressing critical expression that might impair government function.²⁴⁰ The political nature of the Act was obvious from its exclusive use against editors of Democratic-Republican newspapers, and criticism of the resulting trials contributed to the defeat of the Federalists in the 1800 election and the subsequent expiry of the Act.²⁴¹ This example illustrates the peril in mistaking constitutional means for censorial ends: whether the restraint came prior or subsequent to the expression would have meant little to the publishers facing sedition charges. Between the expiration of the Sedition Act and the First World War, virtually no further attempts were made by the U.S. federal government to directly restrict the press apart from during the American Civil War.

system, an aspect of Twitter’s content apparatus that prioritized verified users’ content, as a “lords & peasants system”) and Lauren Sforza, ‘Blue check chaos: Twitter policy spurs confusion and scramble to prove authenticity’ (TheHill.com, 21 April 2023) <<https://thehill.com/policy/technology/3963163-blue-check-chaos-twitter-policy-spurs-confusion-and-scramble-to-prove-authenticity/>> accessed 30 November 2023 (reporting claims by government officials that changes to the verification system would cause harm)

²³⁵ Louw (n 198) 44

²³⁶ A Nichol, G Millar, and A Sharland, *Media Law and Human Rights* (2nd edn, OUP 2009) 7

²³⁷ Roger Fisher, ‘Press Rights vs. Press Responsibilities: The Issues of the First Amendment’ in William L. Rivers & Michael J. Nyhan (eds), *Aspen Notebook on Government and the Media* (Praeger London 1973) 51, 54-55

²³⁸ The language of the American Bill of Rights is naturally quite similar to the English Bill of Rights 1689, likewise expressing rights as negative restraints on the actions of the sovereign

²³⁹ U.S. National Archives and Records Administration, ‘Alien and Sedition Acts (1798),’ (2022) Milestone Documents <<https://www.archives.gov/milestone-documents/alien-and-sedition-acts>> accessed 6 March 2022

²⁴⁰ Nichol, Millar, and Sharland (n 236) 6-7 (discussing *R v Turchin* (1704))

²⁴¹ U.S. National Archives (n 239)

Modern Responses to Media Influence: Access Provision, Marketization

The interrelatedness between law and communications technology is often quite clear in the context of information distribution, particularly from the early modern period onward. The rudiments of information distribution have long featured a certain level of government involvement to realize its public benefit.²⁴² Henry VIII's first Master of the Posts, Brian Tuke, instituted a formal postal network across England for use by the royal court in the early sixteenth century, and Charles I opened the service to the public in 1635.²⁴³ Central organization of the postal services in the American colonies was established by Crown grant at the end of the seventeenth century.²⁴⁴ During Benjamin Franklin's tenure as a Postmaster General for the Crown, the colonial posts surveyed and marked roads, improved routes, and even ran an overnight service between Philadelphia and New York by post rider.²⁴⁵ The American colonists anticipated the importance of a postal service spanning the length of the colonies, even before the onset of the Revolutionary War. Goddard's Constitutional Post, for example, operated 30 offices between the Virginia and New Hampshire colonies and competed with the Crown's service on promises of security and privacy.²⁴⁶

The U.S. Congress periodically adjusted its postal laws to protect the federal service from private competitors who only serviced the most profitable delivery areas.²⁴⁷ British governments in the 19th Century also took steps to improve the distribution of news, rolling back existing censorial measures and gradually removing the Stamp Taxes on newspapers and advertising.²⁴⁸ The Industrial Revolution would introduce further innovations in distribution technology that facilitated the expansion of news media production. The initial cost and difficulty of mail transport was decreased by the establishment of rail and steamboat lines, opening up competition in the delivery business and new markets for print content. Electric telegraphy began to catch on as railway travel and speed had increased (as had accidents and safety concerns).²⁴⁹ The telegraph lines along the railways were adapted for other uses "almost immediately" upon their introduction: "In 1840, the first telegram to excite London, that the Queen had given birth . . . , was carried from Windsor on the Great Western Railway's telegraph line," and in the U.S., telegraph wire between Washington, D.C.

²⁴² See generally William Lewins, *Her Majesty's Mails: A History of the Post-Office and an Industrial Account of its Present Condition* (2nd edn, Sampson Low 1865), ch. 1

²⁴³ Royal Mail Group, 'Our story' (2022) <<https://www.royalmailgroup.com/en/about-us/our-story/>> accessed 6 March 2022

²⁴⁴ U.S. Postal Service, *The United States Postal Service: An American History* (Washington, D.C. 2020) 2

²⁴⁵ *ibid* 2-3

²⁴⁶ *ibid* 3

²⁴⁷ *ibid* 11

²⁴⁸ Jonathan Silberstein-Loeb, 'The Structure of the News Market in Britain, 1870–1914' (Winter 2009) 83 *Business History Review* 759

²⁴⁹ Winston (n 197) 23 ("Early railways were single-track affairs which necessitated, for the first time, instantaneous signaling methods.")

and Baltimore carried inventor Samuel Morse's first public message in 1844.²⁵⁰ Winston notes that "new uses [for the telegraph] were immediately discovered and rapidly developed by the stock market, and especially in North America, newspapers."²⁵¹

Britain and the U.S. would take markedly different approaches regarding comparable aims in realizing the telegraph's public benefit, however, with the British nationalizing the technology under the Post Office to administer public access.²⁵² Governments in Europe like the French, which viewed the telegraph as a substitute for their military communications technologies, simply nationalized Morse's apparatus without necessarily expanding its public availability.²⁵³ But in the United States, the "transmission of intelligence was to be neither necessarily nor exclusively a government function."²⁵⁴ Despite urging by Morse for the U.S. government to buy him out and supervise its subsequent development, American telegraphy remained in private hands.²⁵⁵ Where telegraph lines ended, government-supported postal services continued to relay newspaper reports to editors further afield.²⁵⁶ Policies like these tend to reflect clear normative aims like universal access to postal services or promotion of the newspaper business. Nevertheless, by 1866, one company—Western Union—would come to dominate the American telegraph system, while the Associated Press, through its collusive relationship with Western Union, would come to dominate the news business.²⁵⁷

As noted in Part One, telegraphy dramatically slashed the time and cost involved in transmitting information over long distances. It put horseback express couriers out of business practically overnight and would remain a key part of the global communications infrastructure throughout the 20th Century.²⁵⁸ Furthermore, the uses for which the telegraph was adapted foreshadowed many activities for which social media platforms are used today, including news gathering:

²⁵⁰ Winston (n 197) 23-24

²⁵¹ *ibid* 27

²⁵² Silberstein-Loeb (n 248) 760-61

²⁵³ Winston (n 197) 27; *cf.* Rachel Craufurd Smith and Damian Tambini, 'Measuring Media Plurality in the United Kingdom: Policy Choices and Regulatory Challenges' (2012) 4(1) *Journal of Media Law* 35, 36-41 (highlighting some of differences that have evolved between US and UK rationales for media market intervention)

²⁵⁴ Winston (n 197) 27-28

²⁵⁵ Daniel J Czitrom, *Media and the American Mind: From Morse to McLuhan* (Univ N Carolina P 2010) 21-22

²⁵⁶ Richard B Kielbowicz, 'News Gathering by Mail in the Age of the Telegraph: Adapting to a New Technology' (January 1987) 28(1) *Technology and Culture* 26, 31-32

²⁵⁷ Czitrom (n 255) 23-24

²⁵⁸ Discussed in Part I, 'Social Media News Use'; see also Shelly Freierman, 'Telegram Falls Silent Stop Era Ends Stop' (*New York Times*, 06 Feb 2006) <<https://www.nytimes.com/2006/02/06/technology/telegram-falls-silent-stop-era-ends-stop.html>> accessed 01 April 2023 (Western Union sent its final telegram on January 27, 2006)

“[Stock] speculators began to see the utility of the telegraph and the beginnings of that process which today allows global simultaneous participation in a multitude of international money and commodity markets Newspapers became avid consumers of telegrams, which had a considerable effect on their contents. No longer, as in the eighteenth century, was it possible to scoop a rival with ‘late intelligences’; now, the telegraph rendered news, like soft fruit, perishable—useless if delayed.”²⁵⁹

Development of facsimile transmission capability²⁶⁰ closely followed after the telegraph. Experimentation continued to build on an expanding scientific knowledge base into the 1870s, producing the telephone²⁶¹ as well as the phonograph and kinetoscope not long after.²⁶² These technologies, too, were rapidly adopted for conveying news stories: “The first use of a telephone for news reporting was on 3 April 1877, when word of one of [Alexander Graham] Bell’s lectures was transmitted from Salem [Massachusetts] to the *Boston Globe*.”²⁶³ Other uses devised in the 1870s and 1880s would be familiar to anyone with a mobile phone today: Live theatre, opera and concert performances, lectures, church services, and all-day news reports could all be heard over the telephone, though subject to the low fidelity and limited bandwidth of the early telephone system.²⁶⁴ By the 1890s, experiments with radio waves had demonstrated the possibility of wireless telegraphy.²⁶⁵

While the social order of the late 19th Century remained relatively undisturbed by these further technological advances (despite the fears of commentators of the time),²⁶⁶ some lasting cultural developments have resulted. Czitrom describes how

“In place of the enormous faith invested in the telegraph by the earliest observers, late nineteenth-century thinkers increasingly identified the telegraph and the modern newspaper as both symptom and cause of the frantic pace of industrial life. The disturbing challenge of the periodical press to classical notions of culture began to elicit troubling doubts about the ultimate cultural import of modern communication.”²⁶⁷

²⁵⁹ Winston (n 197) 27-28

²⁶⁰ *ibid* 28-29

²⁶¹ *ibid* 37-41, 48

²⁶² *ibid* 60, Burke (n 209) 289

²⁶³ Winston (n 197) 53

²⁶⁴ *ibid* 59-60

²⁶⁵ *ibid* 69-70

²⁶⁶ *ibid* 59

²⁶⁷ Czitrom (n 255) 21

Recording technologies that spun off from telephone research ushered in a completely new industry in the early 20th Century that changed when and where and how people enjoyed music, just as the novel had invited a new, intimate relationship to narrative works²⁶⁸ in the seventeenth century: “The gramophone extended the social circumstances in which music could be consumed. As with radio, the next technology to come on line, recorded sound systems privatized the musical experience.”²⁶⁹ The introduction of the instant camera at the beginning of the 20th Century changed the way people related to previously fleeting, ephemeral moments, and ancient concerns about rumour, privacy, and reputation took on new complexity²⁷⁰—issues still relevant in the world of iPhones and social media news use.

The U.S. radio market first was dominated by the British Marconi Company, which would be nationalized after the U.S. entered the First World War and then privatized as the Radio Corporation of America, jointly owned by corporate goliaths Westinghouse, General Electric, AT&T, and United Fruit.²⁷¹ In Britain, wireless telegraphy followed the path of the telegraph and telephone and was given over to the General Post Office, but the emerging broadcast radio market of the early 1920s lead the GPO to promote the development of a single licensee, jointly owned by the technology manufacturers (and thus avoiding a monopoly by Marconi): the British Broadcasting Company, which in turn became the BBC in 1927.²⁷² Cross-ownership of news media was pursued in the U.S. as an avenue both for promoting the newspapers and staving off a potential challenge from the radio medium,²⁷³ although some early commercial radio operations in the U.S. were run by other private enterprises, such as department stores, who sought to leverage their radio stations for promotional opportunities.²⁷⁴ Cross-ownership was not permitted in Britain, and the original British Broadcasting Company license “prohibited the broadcast of any news bulletin prior to 7 p.m., the time of the last edition of the evening papers.”²⁷⁵ As with the telegraph, these political decisions would result in further distinctions between the American and British media systems.

²⁶⁸ Cecile Jagodzinski, *Privacy and Print: Reading and Writing in Seventeenth Century England* (Virginia UP 1999)

²⁶⁹ Winston (n 197) 63

²⁷⁰ Solove (n 205) 105-07

²⁷¹ Winston (n 197) 74-79; United Fruit Company likely was included with the other technology companies due to hands-on experience in South and Central America running its subsidiary, Tropical Radio and Telegraph. See Christina S. Drale, ‘The United Fruit Company and Early Radio Development’ (2010) 17(2) J Radio & Audio Media 195

²⁷² Winston (n 197) 82-83

²⁷³ *ibid* 85-87

²⁷⁴ *ibid* 79-80 (AT&T initially developed the concept of leasing radio broadcast time for commercial messages while part of RCA)

²⁷⁵ *ibid* 85-87

Louw contends that while commercial radio in the U.S. developed its industrial character in pursuit of cost-cutting through network arrangements and standardized listening formats, the British public radio model nevertheless developed similar mass-media characteristics due to the pervasive “corporate industrial logic” of the 20th Century that permeated broader society’s intellectual and meaning-making efforts.²⁷⁶ Radio would also pass on many of its institutional tendencies and practices to the nascent television industry.²⁷⁷ American broadcast networks like the National Broadcasting Company and the Columbia Broadcasting System transferred over business and production competencies—and market dominance—from their radio undertakings to the new medium.²⁷⁸

The development of the telegraph directly resulted in the technologies on which the Internet was eventually built. The connections to remote corners of the world we take for granted in discussing the World Wide Web were made long ago—whether by telegraph wire, shortwave radio, or by the early 1960s, communications satellites.²⁷⁹ Social media depends on these existing connections to be ‘social’, and widespread access to social media owes much to the fall in prices for the related technologies like mobile phones that provide network connectivity. This may not, however, mean that systemic obstacles to public function are a thing of the past. As with the private telegraphs of 19th Century America, the development of social media has been largely left to the industry itself—with market forces presumably left to promote desirable characteristics like civic discourse participation. Yet as discussed in the following pages, the concentration and sophistication of the telegraphy infrastructure resulted in significant obstacles that necessitated further public responses. Lessons from these experiences can provide valuable insights into how public policy might be structured around challenges arising from similar factors in social media systems.

Contemporary Responses to Media Influence: State Regulation, Self-Regulation

In the 20th Century, communications regulation grew as a response to increased industrial concentration and sophistication of communications infrastructure. In the U.S., the economic liberty that gave rise to the telegraph had also permitted its consolidation into too few hands. The people naturally became concerned about how those hands were directing the flow of information the people relied on to govern themselves, and applied their public power to loosen the grip of monopolies like Western Union. Broadcasting technologies had

²⁷⁶ Louw (n 198) 38; comparable to the dominant ‘industrial’ ideology of the mid-century period identified by Boltanski and Chiapello (n 282) 165-66

²⁷⁷ Louw (n 198) 38

²⁷⁸ Winston (n 197) 122

²⁷⁹ Sputnik, which carried a radio transmitter, was launched by the USSR in 1957; Telstar 1, the first communications satellite that could send and receive signals, was developed by American Telephone and Telegraph and launched in 1962; see Nokia Bell Labs, ‘Telstar 1’ (2023) <<https://www.bell-labs.com/about/history/innovation-stories/telstar-1/>> accessed 25 March 2023

even greater potential for social influence, in both reach beyond the telegraph wires and richness of the content they could deliver. The tall task of managing that influence necessitated specialization—on both the side of the broadcasters and of newly empowered regulators like the Federal Communications Commission.²⁸⁰ The management of capital is undertaken by agents on behalf of owners under the auspices of contract law and certain norms of professionalism. The management of social forces by bureaucrats on behalf of the public necessitated new auspices: regulatory law and guiding principles like the ‘public interest’ standard to temper regulators’ decision making. But that division of labour also meant separation from, and to an extent exclusion of, the public. As a result, regulators may fall prey to ‘regulatory capture’—industry insiders have the expertise sufficient to do the regulating, and when incorporated into the regulatory process, can tend to favour the regulated industry (as can expert civil servants, who may leave a regulator to become valued additions to private enterprise).²⁸¹

Highlighting a key cultural inflexion during this period is crucial to understanding the differences in ideology between the ‘industrial’ era of broadcasting and content regulation and the ‘network’ era of social media and content moderation. Boltanski and Chiapello identified important transformations in the values promulgated by the avant-garde of multinational businesses over several decades and traced a major shift in the “spirit of capitalism” which began in the 1980s—a broader theorization of capitalism rather than strictly economic that integrates its subjective justifications.²⁸² They contend that a certain set of arguments articulates the logics that justify capitalism relative to its criticisms—the excitement or liberation that comes with involvement in capitalism, the security it offers, and explanations of its fairness and contribution to the common good.²⁸³ These arguments provide a concretization of the spirit of capitalism that “makes the persons it is addressing more aware of the issues that are really at stake, and offers them action models that they will actually be able to use.”²⁸⁴ Boltanski and Chiapello observed that before the 1980s, the processes of capital accumulation typically took the form of large, integrated firms that relied on technical management and hierarchical control to realize economic scale.²⁸⁵ The justificatory logics of this industrial capitalism emphasized the appeal of climbing the

²⁸⁰ I provide an in-depth examination of the advent of broadcast regulation in the U.S. in **Appendix D**

²⁸¹ e.g., Eric Lipton and John M Broder, ‘Regulator Deferred to Oil Industry on Rig Safety’ (*New York Times*, 7 May 2010) <<https://www.nytimes.com/2010/05/08/us/08agency.html>> accessed 25 April 2023; see generally Ernesto Dal Bó, ‘Regulatory Capture: A Review’ (2006) 22(2) *Oxford Review of Economic Policy* 203 (reviewing theoretical and empirical literature on regulatory capture and evidence on the revolving-door phenomenon)

²⁸² Luc Boltanski and Eve Chiapello, ‘The New Spirit of Capitalism’ (2005) 18 *Int J Polit Cult Soc* 161, 162-63

²⁸³ *ibid* 164

²⁸⁴ *ibid* 163-64

²⁸⁵ *ibid* 165-66

corporate ladder towards more prestigious job positions, the security of a long-term career, and the fairness of meritocracy, credentials, and professionalism.²⁸⁶

This period, beginning from the 1930s, also featured the rise and refinement of related governmental technologies—the regulators who could interface with these large enterprises vis-à-vis the businesses’ role in national economic policies, as well as the welfare state that would support workers and others with a more marginal share in the accumulation process. But from the 1980s onward, Boltanski and Chiapello document a new spirit of capitalism supporting new forms of accumulation: “Management literature during the 1990s offered ideals, propositions for the organization of people, and modes for structuring objects and forms of security—all of which were so different in nature from that which had been on offer during the 1960s that it is difficult not to conclude that capitalism has broadly changed its spirit in the last 30 years.”²⁸⁷ Rather than hierarchical, integrated firms, the firms of the new spirit stress organization by projects and networks and leadership by visionaries rather than accredited professionals.²⁸⁸ The excitement for participants in the new spirit comes from varied opportunities for creativity and innovation rather than steady advancement, security through “employability” in future projects, and fairness through mobility and network building.²⁸⁹

These justifications ring familiar in thinking about the typical Silicon Valley start-ups that grew into the mobile phone apps and social media platforms of the 21st Century. Understanding the logics of network-oriented capitalism helps to explain why the social media firms operate as they do. The accumulative power of firms like Facebook derives from the variety and number of network connections it represents rather than the amount of product it can produce efficiently. Their organizational culture prioritizes the characteristics that support the formation of networks—mobility, availability, and variety²⁹⁰—which are visible in the platforms’ personnel and leadership as well as the breadth of services they offer to users.

This understanding also helps frame public policy in relation to the platforms and in comparison with other media systems. Industrial news media firms like broadcast television channels and newspapers organized under the preceding justificatory logics were amenable to governance methods that co-evolved in the same spirit. Regulators could speak from one hierarchy to another, and expectations about a news firm’s contribution to the public interest

²⁸⁶ Boltanski and Chiapello (n 282) 165-66

²⁸⁷ *ibid* 166

²⁸⁸ *ibid* 165-66, 168-69

²⁸⁹ *ibid* 165-66, 168-69

²⁹⁰ *ibid* 168

could be transmitted downwards through professional norms, career advancement, and accreditation or licensure decisions. Under the new spirit, the organizational flatness of social networks lacks such transmission channels for norms, and its paragons (especially younger tech CEOs) typically reject such notions as inimical to both self-management values and valuable opportunities to expand networks.

Boltanski and Chiapello further describe how the network logic of capitalism has gained in power and prominence during the 1990s relative to the preceding industrial logic.²⁹¹ As noted by Mumford, the transition between phases of social and technical relationship is regularly characterized by overlap and persistence of old forms as new orders come to be established.²⁹² To the extent that network logics have displaced or supplanted industrial logic,²⁹³ public policy faces a challenge in being out of step with the new spirit. The neo-liberal turn in British and American politics has looked to markets as the means to order newly-valuable networks, especially through deregulation, but does not necessarily articulate a vision of the public good beyond accumulation for its own sake. Hints of this mismatch can be seen in the political response to the growth of deregulated cable television networks in the 1990s, as detailed in **Appendix D**. The cable firms wanted to expand by adding cable channels that would attract viewers, while the American government wanted to ensure a place for regulated broadcast channels—the resulting “must-carry” compromise sought to affirm both sides’ commitment to television content in the public interest. Internet communications would receive an even freer policy rein as described in my discussion of Section 230²⁹⁴ in the following pages.²⁹⁵ As a governmental technology, regulation affirms a very different set of values and priorities than those embodied by the social media platforms that grew up in the new spirit.

During the 1990s, debates about the ways in which the Internet should be used produced a series of consequential laws and Supreme Court decisions in the United States that remain so in today’s debates over social media policy. Lessig observed that “about the time the Net was coming into the popular consciousness, a particularly seedy aspect of the Net came into view first.”²⁹⁶ Concern about pornography on the Internet became widespread in 1995, not only because of the rise in numbers of adult users of the Internet, but also the number of children using the Internet and the attention paid to the medium by academics,

²⁹¹ Boltanski and Chiapello (n 282) 170

²⁹² Lewis Mumford, *Technics and Civilization* (7th edn, Routledge & Kegan Paul 1955) 215, 268-69 (describing how we live with a “complicated inheritance” of cultural modes, technological adoptions, and social reactions)

²⁹³ Boltanski and Chiapello (n 282) 171-73

²⁹⁴ 47 U.S. Code § 230 (“§ 230” or “Section 230” throughout)

²⁹⁵ See also Part V, ‘Public Policy and the Public Sphere’ – American Perspectives

²⁹⁶ Lawrence Lessig, *Code: Version 2.0* (Basic Books/Perseus 2006) 249

politicians, and the traditional media.²⁹⁷ The concern, particularly with regard to undesirable influence on children, was not a new one. In Victorian Britain, the ‘penny dreadful’ novel industry had grown amidst falling production costs and higher literacy rates, sparking concern over accessibility to content many found unsuitable for impressionable youths.²⁹⁸ The Obscene Publications Act 1857 reflected the Parliament’s determination to tackle such lurid literature, empowering police constables to “seize obscene material and seek its destruction before magistrates.”²⁹⁹ In the 1980s, the debut of the Sony Walkman allowed a more personal and private—and of course, portable—experience of music on cassette tapes,³⁰⁰ and groups like the Parents’ Music Resource Center sought stricter enforcement of obscenity laws to protect children from increasingly accessible pornographic and violent song lyrics.³⁰¹ Video games³⁰² and television³⁰³ received similar critical attention alongside the Internet, prompting the development of ratings systems to better allow consumers to make choices about what content they encountered. Every parent worried about their child’s ‘screen time’ on social media today shares the same awareness and concern over media influence as people have in years past.

The 1996 Communications Decency Act (“CDA”), passed as part of the 1996 Telecommunications Act, sought to criminalize the “knowing transmission of obscene or indecent messages” to a minor as well as the “knowing sending or displaying of patently offensive messages in a manner that is available to” a minor.³⁰⁴ The CDA was challenged as unconstitutional as soon as it was signed into law, and its restrictions were held to be constitutionally overbroad.³⁰⁵ In *Reno v. American Civil Liberties Union*, the Supreme Court affirmed that the transmission and display provisions of the CDA were impermissibly

²⁹⁷ Lessig (n 296) 249

²⁹⁸ John Springhall, “‘A Life Story for the People’? Edwin J. Brett and the London “Low-Life” Penny Dreadfuls of the 1860s’ (Winter 1990) *Victorian Studies* 223, 224 and 230-32

²⁹⁹ Colin Manchester, ‘Lord Campbell’s Act: England’s First Obscenity Statute’ (1988) 9(2) *J of Legal History* 223

³⁰⁰ Sony, ‘Product & Technology Milestones: Personal Audio’ (2022)

<<https://www.sony.com/en/SonyInfo/CorporateInfo/History/sonyhistory-e.html>> accessed 05 July 2022

³⁰¹ See generally Claude Chastagner, ‘The Parents’ Music Resource Center: From Information to Censorship’ (May 1999) 18(2) *Popular Music* 179 (noting also the coincidence of the PMRC’s effort with the effort to establish the Home Audio Recording Tax, a levy on blank audio tapes)

³⁰² See generally Blake J Harris, ‘Content Rated by: An Oral History of the ESRB’ (2019-2020)

<<https://www.esrb.org/about/>> accessed 05 July 2022 (discussing the formation of the Entertainment Software Rating Board in response to concerns about violence in video games)

³⁰³ See generally TV Parental Guidelines Monitoring Board, ‘Our History’ (2021)

<<http://www.tvguidelines.org/aboutUs.html>> accessed 05 July 2022 (providing a development timeline for the ratings system from the 1996 Telecommunications Act onward, along with the technical integration of the ‘V-Chip’)

³⁰⁴ Telecommunications Act of 1996, 104 PUB. L. 104 (the CDA was contained in Title V of the Act); *Reno v. American Civil Liberties Union* [1997] 521 U.S. 844

³⁰⁵ *Reno v. American Civil Liberties Union* [1997] 521 U.S. 844

overbroad in restricting First Amendment activity on the Internet.³⁰⁶ The Court noted the developmental history of the Internet, the growth of its user base to about 40 million people in the mid-1990s, and its evolving uses in transmitting text, sound, pictures, video, and news.³⁰⁷

The Court also situated the World Wide Web in the broader communications media landscape, which is important for how the jurisprudence of other media applies by comparison:

“Access to most Web pages is freely available, but some allow access only to those who have purchased the right from a commercial provider. The Web is thus comparable, from a readers’ viewpoint, to both a vast library including millions of readily available and indexed publications and a sprawling mall offering goods and services. From the publishers’ point of view, it constitutes a vast platform from which to address and hear from a worldwide audience of millions of readers, viewers, researchers, and buyers. Any person or organization with a computer connected to the Internet can ‘publish’ information.”³⁰⁸

The Court further recognized that this access is encountered more deliberately and selectively than communications received by radio or television.³⁰⁹ That difference was significant in prior cases like *FCC v. Pacifica Foundation* and *Sable Communications of California v. FCC*, and the Court held that the Internet did not exhibit the characteristics of invasiveness or scarcity that had qualified a lower level of First Amendment scrutiny in broadcasting regulation cases.³¹⁰ For the Court in *Reno*, the users of the Internet were more akin to town criers and pamphleteers,³¹¹ evoking the deepest traditions of First Amendment protection for speech and the press.³¹² The provisions of the CDA therefore did not provide the level of precision and narrow construction required when a statute regulates speech on the basis of content.³¹³

³⁰⁶ *Reno*, 521 U.S. at 870-79, 885

³⁰⁷ *ibid* 849-52

³⁰⁸ *ibid* 853

³⁰⁹ *ibid* 854

³¹⁰ *ibid* 868-70

³¹¹ *ibid* 870

³¹² McQuail (n 192) 108 (“[P]rint has benefited most directly from democratic and libertarian trends and long-established principles of free expression.”); see also, *e.g.*, *Lovell v. Griffin* [1938] 303 U.S. 444, 452 (“The press . . . comprehends every sort of publication which affords a vehicle of information and opinion.”)

³¹³ *Reno*, 521 U.S. at 874

The dissent in *Reno* afforded greater discussion to less restrictive alternatives than the primary opinion.³¹⁴ That discussion explored how constitutionality principles in zoning law might be applied to ‘cyberspace’ in the way they are applied to physical space through content ‘tagging’ and screening software.³¹⁵ Lessig’s early exploration of this approach influenced the dissent’s discussion,³¹⁶ and Lessig developed the idea further in his later writings on how to draft a constitutional statute that balanced the legitimate objective of keeping minors away from harmful content without undue burden on adults’ communications.³¹⁷ While the majority opinion in *Reno* expressed scepticism about the feasibility of tagging content to facilitate its screening by users,³¹⁸ tagging and flagging content has become a common part of the portfolio of content management tools on social media platforms.³¹⁹ The Congress quickly attempted to address the concerns raised in *Reno* by passing a second statute, the 1998 Child Online Protection Act, which provided a clearer definition of the proscribed material ‘harmful to minors’ and specified means to demonstrate the restriction of minors’ access.³²⁰ But like the CDA, this statute also was challenged and also failed to convince enough of the Justices that the measure was sufficiently narrowly tailored compared to less restrictive alternatives, particularly as blocking and filtering software had come to be viewed as more effective by the Court by the time the case was decided in the early 2000s.³²¹

One key feature of the CDA was not part of the constitutional challenge in *Reno* and has had a profound and lasting legacy, particularly for social media enterprises—Section 230.³²² Section 230(c) provides “protection for ‘Good Samaritan’ blocking and screening of offensive material,” further specifying that “No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider,”³²³ and that

“No provider or user of an interactive computer service shall be held liable on account of . . . any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene,

³¹⁴ *Reno*, 521 U.S. at 886 (J. O’Connor, concurring in the judgment in part and dissenting in part)

³¹⁵ *ibid* 888-91 (drawing on the Supreme Court’s reasoning in *Ginsberg v. New York* [1968] 390 U.S. 629)

³¹⁶ *ibid* 889-90 (citing Lawrence Lessig, ‘Reading the Constitution in Cyberspace,’ (1996) 45 *Emory L J* 869)

³¹⁷ Lessig (n 296) 251-61

³¹⁸ *Reno*, 521 U.S. at 862, 879 & 881

³¹⁹ See, e.g., Kate Crawford and Tarleton Gillespie, ‘What is a flag for? Social media reporting tools and the vocabulary of complaint’ (2016) 18(3) *new media & society* 410

³²⁰ *Ashcroft v. American Civil Liberties Union* [2004] 542 U.S. 656, 660-61

³²¹ *ibid* 666-70

³²² 47 U.S. CODE § 230 (“§ 230” or “Section 230” throughout)

³²³ § 230(c)(1)

lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected.”³²⁴

These provisions grant to Internet service providers and intermediaries³²⁵ so-called “safe harbors against any liability for harmful material their users might provide . . . they are not ‘publishers’ of their users’ content, in the legal sense,” as described by Gillespie.³²⁶

Klonick’s account of the history and development of the statute points to two trial court decisions reached in New York in the years before the passage of the CDA, one in state court and one in federal court, which heavily influenced the Congressional decision to enact Section 230: in the first case, an internet service provider was held not liable for defamatory content because it exercised no user content review policy; while in the second, another internet service provider *was* held liable as a publisher for defamation because it employed automatic screening and blocking software and had content guidelines.³²⁷ Beyond wanting to protect intermediaries from “a wide and unpredictable range of tort liability” that disincentivized content moderation,³²⁸ the Congressional Record shows that the drafters of Section 230 wanted to do far more than avoid intermediaries supposedly being whipsawed by clashing legal obligations: “We can go much further . . . than blocking obscenity or indecency, whatever that means in its loose interpretations. We can keep away from our children things not only prohibited by law, but prohibited by parents. That is where we should be headed,” explained Representative Cox in his presentation of the amendment that added Section 230 to the CDA.³²⁹ Cox continued:

[O]ur amendment will do two basic things: First, it will protect computer Good Samaritans, online service providers, anyone who provides a front end to the Internet, let us say, who takes steps to screen indecency and offensive material for their customers. . . . Second, it will establish as the policy of the United States that we do not wish to have content regulation by the Federal Government of what is on the Internet.”³³⁰

The Section 230 drafters’ vision of Internet regulation diverges sharply from that of the Supreme Court as would be expressed just months later in the *Reno* decision, where the

³²⁴ § 230(c)(2)(A)

³²⁵ § 230(f)(2) (defining an ‘interactive computer service’)

³²⁶ Tarleton Gillespie, ‘Governance of and by platforms’ in Jean Burgess, Alice Marwick, and Thomas Poell (eds), *The SAGE Handbook of Social Media* (SAGE 2018)

³²⁷ Kate Klonick, ‘The New Governors: The People, Rules, and Processes Governing Online Speech’ (2018) 131 Harv L Rev 1598, 1604-05 (discussing *Cubby, Inc. v. CompuServe, Inc.*, 776 F.Supp 135 (S.D.N.Y. 1991) and *Stratton Oakmont, Inc. v. Prodigy Services Co.*, 1995 WL 323710 (N.Y. Sup. Ct. May 24, 1995))

³²⁸ *ibid* 1605

³²⁹ Congressional Record, 141 CONG. REC. H8468-70 (August 4, 1995)

³³⁰ 141 CONG. REC. H8470

Justices clearly understood the Internet to be a cognizable form of communications media with which the government might interact, subject to discernible constitutional limits.

On the apparent basis of a single New York state trial court decision, Section 230 disrupted the development of the sort of jurisprudence for the Internet as had arisen for other network industries connecting “everyone to everyone” like the mail, telegraph, and telephone—common carriers which, “whether used for personal or business purposes, were regarded as public facilities, but for private use.”³³¹ Rather than encourage judges to make distinctions about the limits of content regulation, decide where the public interest could be considered, and establish “criteria of misuse and deliberate harm to others,”³³² Section 230 soon led the Fourth Circuit Court of Appeals to read the scope of the statute broadly, as noted by Klonick.³³³ In a case involving a plaintiff who wanted to hold an Internet service provider liable for defamatory statements posted on the service by a third party, the Fourth Circuit found the provider “immune under § 230 and held that the section precluded not only strict liability for publishers but also intermediary liability for distributors such as website operators. This holding also extinguished notice liability for online intermediaries.”³³⁴

Klonick contends that the appellate court’s analysis of Section 230 “truly shaped the doctrine” concerning intermediaries like the social media platforms that would arise in the following years: “while the court reasoned that § 230 lifted the ‘specter of tort liability’ that might ‘deter service providers from blocking and screening offensive material,’ it also found it was Congress’s design to immunize intermediaries *from any requirement to do so*.”³³⁵ Section 230 thus creates a “paradox” in its applications to both “(1) *encourage* platforms to remove certain kinds of content, and (2) *avoid* the haphazard removal of certain kinds of content,” without any clear prioritization of the intermediaries’ or of the users’ speech interests.³³⁶ This ambiguity regarding “the role of ‘gatekeepers’ such as Internet service providers and search engines, which have a[n] interest in controlling the flow of content, but meet a potential danger when they are treated as publishers and held responsible for content they disseminate,”³³⁷ thus affords Internet intermediaries a significant advantage over

³³¹ McQuail (n 192) 109

³³² *ibid*

³³³ Klonick (n 327) 1606

³³⁴ *ibid*

³³⁵ *ibid* (emphasis added); this is consistent with Rep. Wyden’s statement at the presentation of the amendment, where he relayed the sentiment that the Internet should be governed solely by the “private sector” and that parents could easily rely on “reasonably priced software” from the “neighborhood computer store” to “child-proof the family computer” from online material that makes middle-aged parents “cringe,” 141 CONG. REC. H8470

³³⁶ Klonick (n 327) 1608 (emphasis original)

³³⁷ McQuail (n 192) 38-39

publishers and editors in other media, who cannot simply adopt or shed such responsibility at will.

So far in this Part, I have discussed some of the history and characteristics of the communications technologies that support the creation and circulation of information. These technologies have been adapted to certain social and cultural functions, shaped in part by their concomitant legal treatment:

- Notions of authorship and accountability measures like official licensure developed as mechanical printing became widespread;
- State and private ownership strategies developed as media distribution became concentrated and accessibility became a concern;
- Public and self-regulation developed as media operations demanded more expertise.

Print news functioned to create space in Western societies for the exchange and development of ideas like democracy and personal liberty that were historically transformational. Postal services and telegraphy supported this function to the extent they were readily accessible by the public, and government policies have had a direct hand in extending that access to varying degrees. Like print media did initially, social media creates a space where knowledge about the world can be built and tested by and amongst individuals.

This aspect of Internet communications has long been viewed as among its most virtuous, but as noted in Part One, social media has exhibited a mixed record in fulfilling the expectations of early Internet theorists who envisioned a blossoming of democratic culture around the turn of the millennium as online infrastructure spread.³³⁸ Political activists have adapted traditional methods of engagement to incorporate online media, including the use of news in politics. Social media is viewed as an important venue for political news producers, including alternative and independent media producers often sought out by consumers online. But given that neither social media nor traditional media appear to be displacing the other as trustworthy news outlets (rather, that news consumption overall is declining),³³⁹ should social media be viewed as the herald of a radically new technological age—or is its distinctiveness owed to its particular legal treatment?

³³⁸ Part I, 'Social Media News Use'

³³⁹ Part I, 'Social Media News Use'; e.g., Laura Hazard Owen, 'As traditional news use declines, online news isn't making up the gap' (*NiemanLab*, 14 June 2022) <<https://www.niemanlab.org/2022/06/as-traditional-news-use-declines-online-news-isnt-making-up-the-gap/>> accessed 28 July 2022

Looking Ahead: Continuities and Discontinuities

Whether we view social media as more or less like other communications technologies and industries influences what resources and thinking we can draw on to better understand it. Winston contends that “the received understanding of our current technological situation, the view that we are living in the midst of an ‘Information Revolution’ or at the start of an ‘Information Age’, can be seen rather differently if the histories of the technologies involved are considered.”³⁴⁰ Winston goes on to argue that the supposedly rapid pace of change in communications technology belies the “fundamental continuity” displayed in the West over the past 300 years. The crux of his position is that the histories of electronic communication systems actually show a regular pattern of innovation and diffusion being accommodated by pre-existing social formations.³⁴¹ Winston points out further social necessities which arise from other innovations that serve as important drivers of communications technologies, such as the development of railways and steam-powered ships, along with structural and commercial factors that affect consumer demand.³⁴² On a deeper level, these technologies have “both shaped and been shaped by the diverse processes of social, economic, and political life.”³⁴³

Winston’s perspective does not suggest, however, that we do not live in interesting times. Rather, the key idea is that technological developments catch on after or amidst other social developments that make the utility and implications of the technologies more manifest. This relationship appears regularly in the diffusion of the technologies discussed in this section and invites additional analyses of institutions, economics, and power—dimensions which are also vital to understanding law systems according to their social context and function. Mumford likewise has extensively detailed the “ideological and social preparation” that preceded the historical adoption of industrial processes and mechanization by cultures that were ready to use them.³⁴⁴ Economists such as Read have emphasized the interdependencies and contingencies that had to develop in order to produce even simple modern technologies like the wooden pencil: logging and graphite mining, trade and shipping between distant parts of the world, machining of the metal ferrule, synthesis of the chemicals that make up the eraser, and even the growing of the coffee the workers drink

³⁴⁰ Winston (n 197) 1

³⁴¹ *ibid* 2

³⁴² *ibid* 8-9 (for example, rival inventors Bell and Gray both brought the telephone to market nearly simultaneously as each sought to respond to the ‘social necessity’ present in the circumstances of the 1870s)

³⁴³ *ibid* 9-10; see also James Curran, ‘Media and the Making of British Society, c. 1700-2000’ (2002) 8(2) *Media History* 135 (warning against viewing new communications technology “as an autonomous cause of change” and emphasizing the influence of social context)

³⁴⁴ Mumford (n 292) 4

during the process.³⁴⁵ Contemporary legal scholarship on communications technology in particular has highlighted the importance of sustaining law's normative effects as we adopt additional ways of using cyberspace,³⁴⁶ as well as the role of bureaucracy and private power in shaping how we interact with the Internet.³⁴⁷

By contrast, perspectives which put the technology ahead of the 'social' tend to emphasize characteristics of the technology itself.³⁴⁸ Mid-20th Century media studies, or more specifically *medium* studies, tended to focus on the forms by which people communicate rather than specific messages.³⁴⁹ This approach has produced some insightful observations about the communication process and the consequences of its technological development: Writing in the 1960s, McLuhan identified the potential for television, computers, and satellite communication to globalize the shaping of cultures and societies.³⁵⁰ Interestingly, McLuhan also presaged the use of the same technology for pervasive social control and elections manipulation:

"There's nothing at all difficult about putting computers in the position where they will be able to conduct carefully orchestrated programming of the sensory life of whole populations. I know it sounds rather science-fictional, but if you understood cybernetics you'd realize we could do it today. The computer could program the media to determine the given messages a people should hear in terms of their over-all needs, creating a total media experience absorbed and patterned by all the senses. We could program five hours less of TV in Italy to promote the reading of newspapers during an election, or lay on an additional 25 hours of TV in Venezuela to cool down the tribal temperature raised by radio the preceding month. By such orchestrated interplay of all media, whole cultures could now be programmed in order to improve and stabilize their emotional climate"³⁵¹

While McLuhan nevertheless evinced a blithe optimism about managing the "global village" from its technological "city" centre,³⁵² others like Graeber flatly reject the notion that "we are living in a world created by computers" or that globalization was an "inevitable effect of

³⁴⁵ Leonard E Read, '*I, Pencil: My Family Tree' as told to Leonard E. Read* (Foundation for Economic Education, 1958); Mumford (n 292) 110 (Mumford similarly relates the fountain pen to its broader milieu)

³⁴⁶ Chris Reed, *Making Laws for Cyberspace* (Oxford UP 2012)

³⁴⁷ David Graeber, *The Utopia of Rules: On Technology, Stupidity, and the Secret Joys of Bureaucracy* (Melville House 2015)

³⁴⁸ Daithí Mac Síthigh, *Medium Law* (Routledge 2018) 14

³⁴⁹ *ibid* 13

³⁵⁰ Marshall McLuhan, *Letters of Marshall McLuhan* (Oxford UP 1987) 278

³⁵¹ Eric Norden and Marshall McLuhan, 'Playboy Interview: Marshall McLuhan—A Candid Conversation with the High Priest of Popcult and Metaphysician of Media' (1969) 16(3) *Playboy* 53, 72-74

³⁵² McLuhan (n 350) 278, Norden and McLuhan (n 351) 72-74

technological development, rather than of social and political forces.”³⁵³ Analysing these deeper dimensions in turn focuses on important normative questions that may be elided by a purely deterministic perspective on technology.

Past, present, and future advances in communications technologies and law therefore need a way to be assessed in the broader context of their evolution, including the economics of their development or adoption and the cultural expectations concerning their use or misuse. Burke presents an epistemic framework in which changes or advancements in knowledge may be considered in relation to the changes they bring about in our culture—

“[N]ovel systems . . . either oust or coexist with the structures and attitudes held prior to the change. Our modern view is thus a mixture of present knowledge and past viewpoints which have stood the test of time, and for one reason or another, remain valuable in new circumstances.”³⁵⁴

Adopting this framework allows the relative impact of social media on our lives to be assessed through comparisons with earlier communications technologies like the printing press to see, in terms of culture, what has stuck around and what has been replaced as knowledge has advanced. In turn, we can check whether various proposed means for shaping the public role of social media (like regulation) have fallen out of step with our current circumstances.

Common dimensions between newspapers and social media as media systems suggest where particular analogies and contrasts between the two may be drawn. Hallin and Mancini examined mass circulation newspapers across European countries and developed a four-part framework for comparing media systems, including levels of press circulation, political parallelism, journalistic professionalism, and state intervention.³⁵⁵ They determined that various aspects of market structure that could explain why some countries developed high levels of newspaper reading (as in northern Europe) and others did not (as in southern Europe).³⁵⁶ Further variables like geography and legal tradition have contributed to divergence between British and American media markets, including differences in the extent and form of state involvement, even though both countries exhibit broadly ‘liberal’ media systems.³⁵⁷ Shared differences across this array of variables then tend to distinguish liberal

³⁵³ Graeber (n 347) 30-32

³⁵⁴ Burke (n 209) 11

³⁵⁵ Daniel C Hallin and Paolo Mancini, *Comparing Media Systems: Three Models of Media and Politics* (Cambridge UP 2004) 21

³⁵⁶ *ibid* 22-23

³⁵⁷ *ibid* 41-44; Hardy (n 191) 19

media systems from more polarized media systems.³⁵⁸ Lessig has pointed out the complex web of law, norms, architecture (i.e., computer code), and market forces constraining and supporting free speech in cyberspace³⁵⁹ that similarly characterise social media as a system.

Social media can be taken as exhibiting similarly high levels of circulation for news content as newspapers have in Britain or the U.S.—with an important caveat. Reaching the broad user base of the social media platforms and benefitting from the low cost of such distribution depends on whether the content avoids moderation by the platforms. While our newspapers may have a particular (and even obvious) liberal or conservative leaning, political parties participate directly in online news discourse, posting their own accounts of political events through Twitter and YouTube. Social media also features a wider range of journalistic professionalism compared to traditional newspapers. As discussed in Part One, part of social media’s appeal has been the greater access it affords to amateur citizen-journalists who can provide primary details on distant events. Nevertheless, traditional mass media has shown continued influence online,³⁶⁰ and as discussed in the next section, fact-checking and other moderation practices tend to favour established media outlets over independent journalists, even when the established players get the facts wrong.³⁶¹ Characterizing state involvement in social media demands careful consideration in light of Section 230: despite being less directly ‘regulated’ than licensed media, social media platforms have proven highly responsive to government policy aims, particularly in the wake of the public scrutiny surrounding the Cambridge Analytica story.³⁶² The level of involvement in social media platform content policy by government security services is especially worth consideration, just as with newspapers in years past.³⁶³

Whether strictly in terms of the characteristics of the technology itself or as part of a broader media system with a particular civic-public function, receiving news over social media can be viewed as another step forward from the circumstance of print newspapers being made more widely available by electronic transmission. The printing press and distribution improvements let us read daily news from around the world merely by going to

³⁵⁸ Hardy (n 191) 17-18

³⁵⁹ Lessig (n 296) 235-36

³⁶⁰ Jack M Balkin, ‘Digital speech and democratic culture: A theory of freedom of expression for the information society’ (2004) 79 NYU L Rev 1, 6-10

³⁶¹ Discussed further in Part II, ‘Platform Policies and Technical Capabilities’ – Evaluating Content Moderation at Scale

³⁶² Discussed further in Part II, ‘Platform Policies and Technical Capabilities’ – Facebook

³⁶³ See, e.g., *The New York Times*, ‘Worldwide Propaganda Network Built by the C.I.A.’ (*New York Times*, 26 December 1977) <<https://www.nytimes.com/1977/12/26/archives/worldwide-propaganda-network-built-by-the-cia-a-worldwide-network.html>> accessed 28 July 2022; James Oliver and Nicholas Gilby, ‘Revealed: the secret British plan to keep Italy’s communists from power’ (*The Guardian*, 02 October 2022) <<https://www.theguardian.com/world/2022/oct/02/revealed-the-secret-british-plan-to-keep-italys-communists-from-power>> accessed 29 October 2022

a street corner newsagent. We can now receive articles from the Daily Mail or the New York Times on our phones and immediately publish (via social media) an open letter to the editor (and all our friends and followers) expressing our thoughts about the articles. McLuhan would hardly be surprised by the global technical infrastructure of the social media platforms,³⁶⁴ but how we structure our social interactions—in our laws, institutions, and normative expectations—has substantially shifted from McLuhan’s day.

Also characteristic of social media is the individuated experience it provides in the user’s consumption or production of news content. Jagodzinski points out that amidst the broad public deployment of print to win political and religious battles in the seventeenth century, “increased opportunities for literacy and the emphasis by religious leaders on the development of a personal spirituality created, quite unexpectedly, a greater sense of personal autonomy and a greater acceptance of individual, rather than official, interpretations of text.”³⁶⁵ The increased availability of printed books allowed for many public occasions of communication (the priest reading scripture aloud to his flock in church, the town crier proclaiming some bit of news, the stage performance of a play) to move into private and household spaces and in the ultimate sense of “privatization,” “within the individual reader.”³⁶⁶ The high availability of news on social media and the opportunities it provides for personal interpretation in turn ties in with many of the psychological concepts discussed below in Part Three, including biased assimilation and other phenomena related to social groups.³⁶⁷

The interplay between the content created by a producer and how it is interpreted (and by whom) presents an enduring question about appropriate media system function: the extent to which authors or publishers should be held to account for how their audiences react. Social media exhibits a kind of dual character which both extends this individuated experience (particularly of narratives) and also allows for public, social, and civic uses of news information and discussion, complicating the question of accountability.³⁶⁸ Authorship and audience may be less certain on social media than in newspapers, while social media’s news producers may be less subject to professional norms than journalists. And as covered in the next section’s examination of the moderation policies and automated processes intimate to its distribution of news content, social media is a profoundly mediated

³⁶⁴ Norden and McLuhan (n 351) (“Computer technology can—and doubtlessly will—program entire environments to fulfill the social needs and sensory preferences of communities and nations. The content of that programming, however, depends on the nature of future societies—but that is in our own hands.”)

³⁶⁵ Jagodzinski (n 268) 131

³⁶⁶ *ibid* 49-50

³⁶⁷ Part III, ‘The Social Psychology of Political News and Social Media’

³⁶⁸ Part I, ‘Social Media News Use’ and Part I, ‘The Fake News Problem’

experience.³⁶⁹ Unlike newspapers, where an editor is accountable for deciding the stories a paper prints, social media editorial decisions are made automatically and invisibly by proprietary computer algorithms. More like with the telegraph than with newspapers, this mediation is highly technical, capital intensive, and prone to centralization and consolidation—which likewise makes social media vulnerable to similar misuses and abuses.³⁷⁰ Recognition of these similarities is apparent in some of the American States’ social media ‘consumer protection’ proposals covered in Part Five.

These aspects of social media function and dysfunction can be traced to distinct structural elements that comprise the social media system. Louw suggests that the Internet has exhibited a split potential, both for “de-industrializing and de-institutionalizing communication flows” but also for serving as “another site for institutionalized communication” and the narrowing of discourse.³⁷¹ With regard to social media as a news-media system in particular, it features de-industrialized production and consumption (content made and viewed by individual users) and hyper-industrialized distribution (centralized and automated by the platforms). The implication here is that as compared to industrialized mass-circulation newspapers, the likely ‘sticking point’ in terms of achieving authentic online discourse with a positive social function lies not with user content itself but with the agglomerated power of the social media platforms over its dissemination.

In the next section, I discuss some of the general principles and techniques of content moderation at industrial scale, as well as specific moderation policies and practices at three of the biggest social media platforms, YouTube, Twitter, and Facebook. That discussion provides an in-depth look at the relationship between the technical systems and editorial policies at those platforms—a key dimension where they differ from past news media systems in terms of industrial concentration. This Part concludes with a discussion of research on the efficacy of existing implementations of fake news moderation techniques.

Platform Policies and Technical Capabilities

This section of Part Two focuses on the technical capabilities employed by the social media platforms to moderate content, and also examines the specific policies and strategies of YouTube, Twitter, and Facebook. Because of the scale of their operations, these social media giants rely on automation to regulate the distribution of user generated content—the algorithms that select and order what gets shown to users. To power these algorithms, each

³⁶⁹ Part II, ‘Platform Policies and Technical Capabilities’

³⁷⁰ Winston points out that broadcasting technologies, which differ structurally from the telephone system in allowing a broadcaster to talk to many others at once, also exhibit high centralization and control along with low accessibility; Winston (n 197) 60

³⁷¹ Louw (n 198) 39

of the platforms has developed systematic means for collecting data about their users and about the tweets, posts, videos, and comments their users produce. This section concludes with a discussion about the outcomes of various technological approaches to regulating fake news content. Multiple studies have looked at the efficacy of content moderation technologies made possible by ‘big data’, from network traffic analysis and fake news detection software to natural language processing and machine learning. Other research has scrutinized the use of ‘fact-checking’ techniques by the platforms and third parties. These policies and technical capabilities represent important systemic factors in the status quo flow of news information and interaction on the social media platforms.

Across social media platforms, the ability to implement policy goals—whether internally determined or derived from governmental directives or codes of practice—depends on the technical capabilities of the platforms to engineer their users’ experiences. As noted by Gillespie, the platforms “don’t make the content, but they make important choices about that content: what they will distribute and to whom, how they will connect users and broker their interactions, and what they will refuse,”³⁷² what Facebook CEO Mark Zuckerberg calls the “editorial decisions” he has to make about the scope of content on the platform.³⁷³ The material in this section establishes the general principles and constraints under which the social media platforms perform content moderation. The ways in which their algorithms turn data into action concerning content (and how effectively they can do so regarding fake news) informed how I shaped my empirical research in the following Parts, as well as how I assessed both the status quo of content policy and nascent efforts to regulate or co-regulate the platforms in Part Five.

A fundamental, broadly employed technique, often called the terms of use or terms of service, provides a contractual or private-law basis for the initiation, maintenance, and adjustment (or termination) of the user experience on social media platforms. These agreements also commit users to following the content policies or guidelines promulgated by the companies.³⁷⁴ However, the terms of such agreements are typically not open to negotiation, and individual users have little bargaining power against large global platforms.³⁷⁵ Despite this imbalanced relationship, the users’ acceptance of the terms of service forms the basis of the platforms’ subsequent application of any other technologies at

³⁷² Gillespie (n 326) 1

³⁷³ *The Joe Rogan Experience*, ‘#1863 – Mark Zuckerberg’ (*Spotify*, 25 August 2022) <<https://open.spotify.com/episode/51gxrAActH18RGhKNza598?si=a61ababed4584e76>> accessed 28 August 2022

³⁷⁴ Discussed further *infra* re: YouTube, Twitter, and Facebook

³⁷⁵ Ellen Wauters, Eva Lievens, and Peggy Valcke, ‘Social Networking Sites’ Terms of Use: Addressing Imbalances in the User-Provider Relationship through Ex Ante and Ex Post Mechanisms’ (2014) 5 *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 139

their disposal to monitor and moderate online content. Content moderation in particular may engage a number of different strategic approaches and technical methods. Some platforms, typically at a smaller scale like blogging platform Medium, may follow an ‘artisanal’ strategy, employing in-house moderators to make decisions about content.³⁷⁶ This approach usually exhibits little or no use of automation, allowing flexibility and responsiveness to context—but potentially requiring “significant organizational costs” as companies scale up to a larger number of users and wider variety of content.³⁷⁷

Caplan notes that firms face additional pressure to formalize moderation structures where they become subject to legal rules (like the *NetzDG* law in Germany or the Online Safety Bill and other measures discussed in Part Five) that require specific procedures be followed in a short time frame with regard to various types of prohibited speech.³⁷⁸ Other platforms may rely on an active user community to provide volunteer content moderation services, as on sites like Wikipedia and Reddit, similar to the ways early Internet bulletin boards and forums have done historically,³⁷⁹ authorising administrators to make content decisions, but distributed across large user communities.³⁸⁰ The ‘community-reliant’ approach thus puts content policy decisions more directly in the hands of the users, which necessarily means that platforms taking this approach also trade off some level of control over content policies and risk coming into conflict with the volunteer base where such platforms attempt to impose specific rules.³⁸¹ Facebook, Twitter, and YouTube instead rely on an ‘industrial’ approach to apply their content policies to very large user bases that generate huge amounts of content.³⁸² By ‘operationalizing’ moderation decisions—breaking the processes down into repeatable applications of formal rules—these platforms allow for routine action upon the detection of prohibited content.³⁸³ Such detection relies on a combination of user flagging and automated technology, particularly as automated detection may be effective at finding some kinds of content like spam or nudity but less effective at correctly identifying hate speech, for instance.³⁸⁴

³⁷⁶ Robyn Caplan, ‘Content or Context Moderation?: Artisanal, Community-Reliant, and Industrial Approaches’ (*Data & Society*, 14 November 2018) 17 <<https://datasociety.net/library/content-or-context-moderation/>> accessed 28 August 2022

³⁷⁷ *ibid* 18-19

³⁷⁸ *ibid* 19

³⁷⁹ Gillespie (n 326) 15

³⁸⁰ Caplan (n 376) 20

³⁸¹ *ibid* 21-22

³⁸² *ibid* 23-24

³⁸³ *ibid*

³⁸⁴ *ibid* 24-25

Social media platforms hosting user-generated content typically provide a mechanism to users for reporting offensive content ('flagging').³⁸⁵ Content flagging by users may engage other human content moderation systems, whether in-house (as in the artisanal or community-reliant approaches described above) or outsourced to third-party firms.³⁸⁶ When these systems are engaged, flagged content enters a review process where a human evaluator makes a judgment whether the content should remain on the platform.³⁸⁷ Platforms can also rely on flagging to operate automated filtering systems, allowing users to avoid certain content they do not wish to see—provided it is correctly flagged.³⁸⁸ Crawford and Gillespie describe flagging as more than merely a “technical feature: It is a complex interplay between users and platform, humans and algorithms, and the social norms and regulatory structures of social media.”³⁸⁹

While ubiquitous, the content flag serves both its practical purpose of generating data necessary to regulate content at industrial scale, and also a symbolic, expressive function for both the platform architects and the platform users relating their concerns about how the platform community should carry on.³⁹⁰ In its expressive function, insofar as it is engaged by the users, content flagging reflects the sentiments of the users who are the “most motivated to complain”—who do so out of a myriad of motivations, values and concerns which remain opaque in the action of flagging.³⁹¹ The reliability of human annotation becomes materially important in the operation of automatic detection systems, forming an “upper bound” on the performance of algorithms designed to evaluate social media content for hate speech, for instance.³⁹² Such algorithms rely on training data—manually annotated content that accurately matches a classification like ‘hate speech’ or ‘misinformation’—to make subsequent yes-or-no decisions about additional content.³⁹³ While the user base of a social media platform may provide an enormous amount of training data through content

³⁸⁵ Crawford and Gillespie (n 319) 411

³⁸⁶ Sarah T Roberts, ‘Commercial Content Moderation: Digital Laborers’ Dirty Work’ (2016) Media Studies Publications, Paper 12 <<https://ir.lib.uwo.ca/commpub/12>> 1-2

³⁸⁷ *ibid*

³⁸⁸ Gillespie (n 326) 19-20

³⁸⁹ Crawford and Gillespie (n 319) 411

³⁹⁰ *ibid* 412

³⁹¹ Gillespie (n 326) 19

³⁹² Bjorn Ross and others, ‘Measuring the Reliability of Hate Speech Annotations: The Case of the European Refugee Crisis’ (2016) 17 Proceedings of the Third Workshop on Natural Language Processing for Computer-Mediated Communication 6

³⁹³ *ibid*

flagging, some platforms also rely on ‘expert’ annotators or specifically trained workers at commercial content moderation firms in order to provide “reliable” classification data.³⁹⁴

Each of the major social media platforms implements various technical and policy strategies to different degrees in order to meet internal and external expectations about permissible user-generated content. Their business models integrate vast amounts of data and content to keep users engaged. In the following sections, I discuss recent strategy developments at YouTube, Twitter, and Facebook that shape the news environment their users experience. My Thesis focuses on these large platforms because of their shared reliance on automated content moderation systems, a factor that distinguishes them not only from broadcasting and newspapers but also from typically smaller social media platforms that rely on artisanal or community-driven approaches to moderation. In addition, YouTube, Twitter, and Facebook have attracted considerable policymaker attention due to their scale and perceived roles in public discourse and civic life.

YouTube

In 2019, YouTube made changes to its content policy to prohibit videos promoting hate speech and supremacist ideology. As a result, a number of history teachers suddenly found their YouTube channels had been automatically removed by the platform: YouTube’s content moderation process did not discriminate between World War II documentary footage for educational purposes and material glorifying the Third Reich.³⁹⁵ While YouTube allowed for appeals and review of the removals, the incident reflects the tension between the breadth of social media content, the pressure on platforms to conform to local laws, the actual possibilities of content moderation, and the business model of social media. YouTube publishes semi-annual transparency reports detailing its efforts to comply with local laws like those in Germany which prohibit certain content.³⁹⁶ Content subject to local laws may tend to run afoul of YouTube’s “Community Guidelines” and thus be removed from the platform globally, but in some cases might require a more specific legal assessment and only local removal.³⁹⁷ The company describes a “‘people + machine’ framework” to monitor complaints and remove content it identifies as prohibited.³⁹⁸ The people involved include

³⁹⁴ Gillespie (n 326) 19; Zeerak Waseem, ‘Are You a Racist or Am I Seeing Things? Annotator Influence on Hate Speech Detection on Twitter’ (2016) Association for Computational Linguistics, Proceedings of 2016 Workshop on Natural Language Processing and Computational Social Science 138

³⁹⁵ Jim Waterson, ‘YouTube blocks history teachers uploading archive videos of Hitler’ (*The Guardian*, 06 June 2019) <<https://www.theguardian.com/technology/2019/jun/06/youtube-blocks-history-teachers-uploading-archive-videos-of-hitler>> accessed 24 August 2022

³⁹⁶ Google, ‘Removals under the Network Enforcement Law’ (*Google.com*, 31 December 2019) <https://transparencyreport.google.com/netzdg/youtube?hl=en_GB> accessed 28 July 2022

³⁹⁷ *ibid*

³⁹⁸ *ibid*

“policy specialists, lawyers, engineers, product managers, data analysts, content reviewers, operations analysts, emerging threat analysts, . . . the global community of users, NGOs, governments and industry partners – including other technology companies.”³⁹⁹ But the amount of content posted to YouTube—over 500 hours of video uploaded each minute, watched at a rate of 1 billion hours per day—seems to belie an overwhelming reliance on technical capability (the ‘machine’ side of the framework) rather than the “‘about 10,000 people that are focused on controversial content’” at the company per YouTube CEO Susan Wojcicki.⁴⁰⁰

The centrepiece of YouTube’s business is advertising sales, made attractive by the success of its algorithmic recommendation engine that keeps users watching more and more videos on the platform.⁴⁰¹ YouTube’s recommendation engine also powers its technical means of managing prohibited content and so-called “borderline” content that doesn’t violate Community Guidelines but “could misinform users in harmful ways.”⁴⁰² Rather than removing borderline content outright, the platform can also exercise the capability to not include certain videos in the over 200 million recommendations it makes each day on users’ homepages.⁴⁰³ Yet this capability nevertheless depends on “human evaluators and experts” to “train the machine learning systems that generate recommendations.”⁴⁰⁴ The Community Guidelines themselves present policy statements in broad terms such as the following, accompanied by generic examples suggesting corresponding violations:

“Content intended to impersonate a person or channel is not allowed on YouTube . . . [such as] Channels impersonating an existing news channel.”⁴⁰⁵

“Don’t post content on YouTube if it fits any of the descriptions below . . . [including] Misattributed content: Content that may pose a serious risk of egregious harm by falsely claiming that old footage from a past event is from a current event.”⁴⁰⁶

³⁹⁹ Google (n 396)

⁴⁰⁰ Lesley Stahl, ‘How does YouTube handle the site’s misinformation, conspiracy theories, and hate?’ (*60 Minutes*, 01 December 2019) <<https://www.cbsnews.com/news/is-youtube-doing-enough-to-fight-hate-speech-and-conspiracy-theories-60-minutes-2019-12-01/>> accessed 28 July 2022

⁴⁰¹ *ibid*

⁴⁰² The YouTube Team, ‘Continuing our work to improve recommendations on YouTube’ (25 January 2019) <<https://blog.youtube/news-and-events/continuing-our-work-to-improve/>> accessed 24 August 2022

⁴⁰³ *ibid*

⁴⁰⁴ *ibid*

⁴⁰⁵ YouTube Help, ‘Impersonation Policy’ (*Google*, 2022) <https://support.google.com/youtube/answer/2801947?hl=en&ref_topic=9282365> accessed 28 August 2022

⁴⁰⁶ YouTube Help, ‘Misinformation policies’ (*Google*, 2022) <<https://support.google.com/youtube/answer/10834785>> accessed 28 August 2022

“Certain types of misleading or deceptive content with serious risk of egregious harm are not allowed on YouTube. This includes certain types of misinformation that cause real-world harm, like certain types of technically manipulated content, and content interfering with democratic processes . . . [such as] Videos that contain hacked info about a political candidate shared with the intent to interfere in an election” or claims about fraud in past elections in the U.S., Germany, and Brazil.⁴⁰⁷

The policies include exceptions where the content “includes additional context in the video, audio, title, or description” or is intended “to condemn, dispute, or satirize misinformation that violates our policies.”⁴⁰⁸ The examples meanwhile include caveats that the examples are not exclusive and admonish users not to “post content if you think it might violate these policies.”

The vagueness of YouTube’s policies, along with the opacity of its enforcement decisions, has led to some noteworthy controversies, particularly where political influence may have played a key role. In early 2020, for example, following the assassination of a prominent Iranian general by a U.S. drone strike in Iraq, YouTube took decisive action against London-based Press TV, the English-language YouTube channel associated with the Islamic Republic of Iran Broadcasting service, deleting the channel for violating YouTube’s policies.⁴⁰⁹ Press TV has claimed that this and similar previous actions were taken without prior notice and may have reflected “ramped up” efforts by then-President Trump and the U.S. State Department to apply pressure to the Iranian government.⁴¹⁰ Press TV’s *presstv.com* Internet domain was subsequently seized by the U.S. government, although its *presstv.co.uk* site remains accessible and redirects users to the *presstv.ir* domain based in Iran.⁴¹¹

Twitter

In 2018, then-CEO of Twitter, Inc. Jack Dorsey testified before the U.S. Congress about his company’s content moderation practices.⁴¹² Dorsey stated that the social media

⁴⁰⁷ YouTube Help, ‘Elections misinformation policies’ (Google, 2022)

<<https://support.google.com/youtube/answer/10835034>> accessed 28 August 2022

⁴⁰⁸ ‘Misinformation policies’ (n 406); ‘Elections misinformation policies’ (n 407)

⁴⁰⁹ Matthew Moore, ‘YouTube deletes Iranian channel Press TV UK for flouting ban’ (*The Times*, 29 January 2020) <<https://www.thetimes.co.uk/article/youtube-deletes-iranian-channel-press-tv-uk-for-flouting-ban-gxz539v80>> accessed 28 August 2022

⁴¹⁰ Press TV, ‘Google again blocks access to YouTube account of Iran’s Press TV’ (19 September 2020) <<https://www.presstv.ir/Detail/2020/09/19/634536/Iran-Press-TV-YouTube-Google>> accessed 28 August 2022

⁴¹¹ See <https://www.presstv.com/>; cf. <https://www.presstv.co.uk> accessed 28 August 2022

⁴¹² US House Committee on Energy and Commerce, *Testimony of Jack Dorsey Chief Executive Officer Twitter, Inc.* (5 September 2018) (hereinafter ‘Dorsey House Testimony’); US Senate Select Committee on

platform's purpose is "to serve the public conversation" and that "Twitter is used as a global town square, where people from around the world come together in an open and free exchange of ideas."⁴¹³ Similar to Facebook's Community Guidelines, Twitter uses policy statements to inform users about what kinds of content may be prohibited. Dorsey detailed the company's specific efforts following the 2016 U.S. elections, including account verification, a Political Campaigning Policy requiring certain disclosures by political advertisers, and partnerships with federal law enforcement agencies, academics, and NGOs.⁴¹⁴ Dorsey also disclaimed the notion that political ideology informed Twitter's moderation decisions, explaining that the company relies on machine learning algorithms to organize the content generated by its 335 million users.⁴¹⁵ Twitter's algorithms in turn are tuned around several factors, described by Dorsey, including content ranking (showing content based on potential interest, rather than in chronological order), search results (showing content based on popularity and on keywords or hashtags), and behavioural signals (showing content based on "how accounts behave and react to one another").⁴¹⁶

Twitter's Civic Integrity Policy sets out the platform's rules on content pertaining to political elections and other civic processes. In 2020, Parag Agrawal, Twitter's CTO (who subsequently succeeded Dorsey as CEO), described the company's approach to misinformation around election participation as follows:

"[W]e care about and we take action on content which might misinform people who say you should vote on November 5th, when election day is November 3rd. And, we do not try to determine uh what's true or false when someone takes a policy position or when someone says the sky is purple or blue, or red for that matter. Our approach for misinformation is also not one that's focused on taking content down as the only measure, which is the regime we all have operated in for many years. But it's an increasingly nuanced approach with a range of interventions, where we think about whether or not certain content should be amplified without context, or whether it's our responsibility to provide some context so that people can see a bunch of information, but also have the ability and ease to discover all the

Intelligence, *Testimony of Jack Dorsey Chief Executive Officer Twitter, Inc.* (5 September 2018) (hereinafter 'Dorsey Senate Testimony')

⁴¹³ Dorsey House Testimony (n 412) 1

⁴¹⁴ Dorsey Senate Testimony (n 412) 5-7

⁴¹⁵ Dorsey House Testimony (n 412) 3; by contrast, Twitter only employed 3,920 full-time employees as of December 31, 2018: Twitter, Inc., *Fiscal Year 2018 Annual Report* (2019) <<https://investor.twitterinc.com/financial-information/annual-reports/default.aspx>> accessed 26 August, 2022

⁴¹⁶ Dorsey House Testimony (n 412) 3-6

conversation and context around it, to inform themselves about what they choose to believe in.”⁴¹⁷

The additional context Twitter provides includes linking to “credible sources” and “additional conversation” about a topic.⁴¹⁸ Agrawal’s civic integrity approach was rolled out by the company ahead of the 2020 U.S. elections, including labels for or removals of “false or misleading information intended to undermine public confidence” in the election, including “disputed claims that could undermine faith in the process itself.”⁴¹⁹

The Civic Integrity Policy likewise provides examples of false or untrue information that generally does not violate the policy, including “organic content that is polarizing, biased, hyperpartisan, or contains controversial viewpoints expressed about elections or politics; [and] discussion of public polling information.”⁴²⁰ Although pseudonymous or parody accounts discussing politics are generally permitted, the Policy does note that action may nevertheless be taken against humorous or satirical Tweets “given the significant risks of confusion about key election information.”⁴²¹ Twitter further sets out consequences for violation of the policy, including an account lockout and “strike” system, required labelling or deletion of prohibited content, and reduced visibility or redistribution of the Tweet.⁴²² The warnings about satirical Tweets and the strike system were added to the policy after the 2020 election.⁴²³

Agrawal has discussed some of the tensions within Twitter’s framework, including its policy around ‘public interest’ content sourced from elected officials or political candidates and its other policies like Civic Integrity.⁴²⁴ That tension was most notably displayed in the 2021 permanent suspension of U.S. President Donald Trump, who made two Tweets which Twitter determined were in violation of its Glorification of Violence policy.⁴²⁵ The company had previously stated that “[b]locking a world leader from Twitter or removing their controversial Tweets would hide important information people should be able to see and

⁴¹⁷ Gideon Lichfield, ‘EmTech Stage: Twitter’s CTO on misinformation’ (*MIT Technology Review*, 18 November 2020) <<https://www.technologyreview.com/2020/11/18/1012066/emtech-stage-twitters-cto-on-misinformation/>> accessed 28 July 2022

⁴¹⁸ *ibid*

⁴¹⁹ Twitter Safety, ‘Expanding our policies to further protect the civic conversation’ (10 September 2020) <https://blog.twitter.com/en_us/topics/company/2020/civic-integrity-policy-update> accessed 26 August 2022

⁴²⁰ Twitter, ‘Civic integrity policy’ (October 2021) <<https://help.twitter.com/en/rules-and-policies/election-integrity-policy>> accessed 28 July 2022

⁴²¹ *ibid*

⁴²² *ibid* (5 or more strikes leading to permanent suspension of the user’s account)

⁴²³ Cf. Twitter, ‘Civic integrity policy’ (October 2020) <on file with author>

⁴²⁴ Lichfield (n 417)

⁴²⁵ Twitter, ‘Permanent suspension of @realDonaldTrump’ (8 January 2021) <https://blog.twitter.com/en_us/topics/company/2020/suspension> accessed 28 July 2022

debate . . . [and] would also not silence that leader, but it would certainly hamper necessary discussion around their words and actions.”⁴²⁶ Twitter also relies on its “partnerships” with law enforcement agencies like the FBI and supranational organizations like NATO to make decisions about whether networks of accounts constitute “state-affiliated information operations” (and subsequently taking action against them, such as permanent suspensions), as it has done in relation to certain users in Armenia, for example.⁴²⁷ Twitter’s policy on blocking links similarly explains that the company relies on “a combination of factors to identify links that we may ultimately take enforcement action against,” including “collaborative information sharing with industry peers and trusted NGO partners.”⁴²⁸

Facebook

Facebook likewise has made major, platform-wide adjustments in the circulation of news content from publishers and businesses (compared to content from friends, family, and user groups),⁴²⁹ albeit on a much larger scale than Twitter’s user base.⁴³⁰ Facebook’s News Feed function provides users with a limited selection of content from Pages that Facebook has determined will spark conversations between users.⁴³¹ Company chief executive Mark Zuckerberg recently described the operation of Facebook’s recommendation engine as relying on posts and comments by a user’s friends and social groups as a “signal” for what to recommend to that particular user, reflecting what Zuckerberg says Facebook users actually wanted from the platform (which was not necessarily news content).⁴³² Zuckerberg stated that a post from a family member announcing a new baby, for instance, shows up on a user’s News Feed not because the algorithm ‘knows’ what a baby announcement means *per se*, but because the post receives positive reactions from and is shared intensively by a mutually connected group of users.⁴³³

However, by prioritizing community content expected to generate traffic on the platform, Facebook necessarily deprioritized third-party news publishers (given the limited

⁴²⁶ Twitter, ‘World Leaders on Twitter’ (05 January 2018) <https://blog.twitter.com/en_us/topics/company/2018/world-leaders-and-twitter> accessed 26 August 2022

⁴²⁷ Twitter Safety, ‘Disclosing networks of state-linked information operations’ (23 February 2021) <https://blog.twitter.com/en_us/topics/company/2021/disclosing-networks-of-state-linked-information-operations> accessed 20 September 2022

⁴²⁸ Twitter, ‘Our approach to blocking links’ (July 2020) <<https://help.twitter.com/en/safety-and-security/phishing-spam-and-malware-links>> accessed 01 November 2022

⁴²⁹ Laura Hazard Owen, ‘Facebook drastically changes News Feed to make it “good for people” (and bad for most publishers)’ (*NiemanLab*, 11 January 2018) <<https://www.niemanlab.org/2018/01/facebook-drastically-changes-news-feed-to-make-it-good-for-people-and-bad-for-most-publishers/>> accessed 28 July 2022

⁴³⁰ Facebook has billions of users compared to Twitter’s millions

⁴³¹ Hazard Owen (n 429)

⁴³² *The Joe Rogan Experience* (n 373)

⁴³³ *ibid*

‘space’ on a user’s Feed).⁴³⁴ The 2018 News Feed change followed a limited roll-out of similar changes for users in a number of Facebook’s national markets, including Bolivia, where a prominent news publisher (*Página Siete*, not to be confused with American celebrity gossip tabloid *Page Six*) observed a drop in traffic coming from Facebook to its website.⁴³⁵ The New York Times suggested that the de-emphasis of “professional” news outlets (including the New York Times itself) by Facebook brooked the “unintended consequences” of promoting fake and sensationalist news stories.⁴³⁶ Presumably, user-shared information (also presumably false) would be advantaged against official, subscription-based news sites, for instance, “forced to spend money to place themselves in the News Feed.”⁴³⁷

Subsequently, serious violent events have been laid at the feet of Facebook’s News Feed changes by professional news outlets, purportedly resulting from Facebook’s exacerbation of (among other instances) ethno-religious conflict in Sri Lanka⁴³⁸ and class struggle in France.⁴³⁹ But Facebook has also made moderation decisions with discrete impacts on free expression itself, including its 2018 shutdown and subsequent reinstatement of leftist South American media network Telesur’s English-language Facebook Page.⁴⁴⁰ That instance highlights a particular challenge for news publishers on the Facebook platform: understanding the content moderation decisions made by the company. Facebook gave Telesur several different answers for why the media outlet had been removed, ostensibly related to a Terms of Use violation: a technical glitch, instability on the platform, or the detection of “suspicious activity” might have been the actual cause of the automatic suspension.⁴⁴¹ Telesur and other publishers are not alone in pondering how they might

⁴³⁴ Hazard Owen (n 429)

⁴³⁵ Sheera Frenkel, Nicholas Casey, and Paul Mozur, ‘In Some Countries, Facebook’s Fiddling Has Magnified Fake News’ (*New York Times*, 14 January 2018) <<https://www.nytimes.com/2018/01/14/technology/facebook-news-feed-changes.html>> accessed 28 July 2022

⁴³⁶ *ibid*

⁴³⁷ *ibid*

⁴³⁸ Amanda Taub and Max Fisher, ‘Where Countries Are Tinderboxes and Facebook Is a Match’ (*New York Times*, 21 April 2018) <<https://www.nytimes.com/2018/04/21/world/asia/facebook-sri-lanka-riots.html>> accessed 28 August 2022

⁴³⁹ Ryan Broderick and Jules Darmanin, ‘The “Yellow Vest” Riots in France Are What Happens When Facebook Gets Involved With Local News’ (*BuzzFeed*, 06 December 2018) <<https://www.buzzfeednews.com/article/ryanhatesthis/france-paris-yellow-jackets-facebook>> accessed 28 July 2022

⁴⁴⁰ Sam Biddle, ‘Facebook suspended a Latin American news network and gave three different reasons why’ (*The Intercept*, 20 August, 2018) <<https://theintercept.com/2018/08/20/facebook-suspended-a-latin-american-news-network-and-gave-three-different-reasons-why/>> accessed 28 July 2022

⁴⁴¹ *ibid*

discern the rules of acceptable content on Facebook, as Facebook's own employees have also expressed frustration with its rulemaking and enforcement process.⁴⁴²

That process, apparently an ad hoc accumulation of guidelines spanning 1,400 pages, reflects the efforts of a relatively small number of Facebook engineers and lawyers to “distill highly complex issues into simple yes-or-no rules” for training the company's thousands of outsourced moderators around the world to evaluate posts.⁴⁴³ As Gillespie has observed, the development of platform rules and enforcement arises over time and often in response to unexpected contentious issues—and not necessarily always according to a coherent concept.⁴⁴⁴ The scale of the platform, as with YouTube and Twitter, means that the vast majority of policy violations are identified by Facebook's “technical systems,” as described by Facebook's head of global policy management Monika Bickert.⁴⁴⁵ Bickert explains that Facebook's approach tries to strike a balance between “nuances to account for every situation” and “a set of policies we can enforce accurately and . . . explain cleanly.”⁴⁴⁶ But achieving that balance demands a certain level of quality in the data about content on the platform in order for Facebook's algorithm to work as intended. Beyond the difficulty of drawing lines between political speech and disinformation, Facebook has only limited quality control over the moderation companies it contracts with to provide the platform with essential data about users' posts.⁴⁴⁷

As discussed in the previous Part, influence has become a key concern among policy makers, not just with regard to advertising but with the content users are generating and sharing on the social media platforms. In the aftermath of the Cambridge Analytica exposures, politicians called for Facebook to provide answers about social media campaigns to influence elections, including the use of fake news and other misinformation.⁴⁴⁸ Facebook's COO, Sheryl Sandberg, responded to criticism of Facebook's political advertisements policy in 2019 by reiterating the company's position that it takes “political ads because we really believe they are part of the political discourse.”⁴⁴⁹ Sandberg has also

⁴⁴² Max Fisher, ‘Inside Facebook's Secret Rulebook for Global Political Speech’ (*New York Times*, 27 December 2018) <<https://www.nytimes.com/2018/12/27/world/facebook-moderators.html>> accessed 28 July 2022

⁴⁴³ *ibid*

⁴⁴⁴ Gillespie (n 326) 13-14

⁴⁴⁵ Fisher (n 442)

⁴⁴⁶ *ibid*

⁴⁴⁷ *ibid*

⁴⁴⁸ US Senate Committee on the Judiciary and US Senate Committee on Commerce, Science and Transportation, *Testimony of Mark Zuckerberg Chairman and Chief Executive Officer, Facebook* (10 April 2018) 1 (“Zuckerberg Senate Testimony”)

⁴⁴⁹ Joan E Solsman and Queenie Wong, ‘Facebook isn't taking political ads “for the money,” COO Sandberg says’ (*CNet.com*, 22 October 2019) <<https://www.cnet.com/news/facebook-isnt-taking-political-ads-for-the-money-coo-sheryl-sandberg-says/>> accessed 28 July 2022

pushed back against the notion of “echo chambers” on Facebook, explaining that “it is unequivocally true that Facebook usage and uses of social media show you broader points of view, not narrower points of view.”⁴⁵⁰

While its data privacy changes aimed to reduce the leveraged reach of analytics firms and app developers,⁴⁵¹ Facebook has also focused on understanding how “bad actors [have] used coordinated networks of fake accounts” in order to combat election interference.⁴⁵² In his Senate testimony, Zuckerberg reported the platform’s deployment of “advanced AI tools” to automate the detection and sanction of inauthentic behaviour designed to game Facebook’s news distribution system, as well as changes made to Facebook’s advertising policy.⁴⁵³ In the months following the U.K. and U.S. hearings in 2018, Facebook also began a partnership with the Digital Forensic Research Lab, a project of the Atlantic Council think-tank that aimed to “independently monitor disinformation” and research the impact of social media on elections.⁴⁵⁴ The partnership opened Facebook to cooperation with “outside experts, governments, and other companies . . . [to help Facebook’s] security, policy and product teams get real-time insights and updates on emerging threats and disinformation campaigns from around the world . . . and ensure that Facebook plays a positive role during elections,” according to a Facebook executive at the time.⁴⁵⁵

But as the 2020 U.S. elections approached, Facebook faced yet more allegations that it had failed to do enough to combat fake accounts and fake content attempting to influence elections in Europe, India, South America, and elsewhere.⁴⁵⁶ Facebook took steps to scrutinize advertisements and paid promotions (so-called ‘boosted posts’)⁴⁵⁷ in advance of a restricted period before then-upcoming elections in the U.S., Germany, and France, and prohibited ads that made claims about corrupt election administration or voter fraud, for

⁴⁵⁰ Solsman and Wong (n 449)

⁴⁵¹ Federal Trade Commission, ‘FTC Imposes \$5 Billion Penalty and Sweeping New Privacy Restrictions on Facebook’ (24 July 2019) <<https://www.ftc.gov/news-events/news/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions-facebook>> accessed 28 August 2022 (penalties including structural changes to the company)

⁴⁵² Zuckerberg Senate Testimony (n 448) 4

⁴⁵³ *ibid* 5-6

⁴⁵⁴ Atlantic Council, ‘Atlantic Council’s Digital Forensic Research Lab Partners with Facebook to Combat Disinformation in Democratic Elections’ (17 May 2018) <<https://www.atlanticcouncil.org/news/press-releases/atlantic-councils-digital-forensic-research-lab-partners-with-facebook/>> accessed 28 July 2022

⁴⁵⁵ *ibid*

⁴⁵⁶ Craig Silverman, Ryan Mac, and Pranav Dixit, “‘I Have Blood on My Hands’: A Whistleblower Says Facebook Ignored Global Political Manipulation” (*BuzzFeed*, 14 September 2020) <<https://www.buzzfeednews.com/article/craigsilverman/facebook-ignore-political-manipulation-whistleblower-memo>> accessed 28 July 2022

⁴⁵⁷ Sara Fischer, ‘Facebook says new pre-election political ad rules apply to boosted posts’ (*Axios*, 27 October 2020) <<https://www.axios.com/2020/10/27/facebook-says-new-pre-election-political-ad-rules-apply-to-boosted-posts>> accessed 28 July 2022

instance.⁴⁵⁸ However, political campaigners claimed that the platform enforced its policies in a manner that disadvantaged their campaigns—which Facebook denied—and critics faulted the company for not being more transparent about policy enforcement decisions.⁴⁵⁹ Facebook also restricted specific user groups, placing the groups on ‘probation’ and requiring the administrators of those groups to pre-screen posts by a group’s users in order to discourage the spread of false claims about the 2020 U.S. election.⁴⁶⁰ Following the election, Facebook launched a new strategy for dealing with misinformation on the platform by providing additional context for content which its fact-checkers have previously evaluated. That strategy includes pop-up messages (warning users before they liked or shared content) that a particular Page had received “false information” ratings by fact-checkers, as well as reducing the News Feed distribution (visibility) of posts from users who repeatedly share such rated content.⁴⁶¹

Evaluating Content Moderation at Scale

In the remainder of this section, I delve deeper into what distinguishes the ‘hyper-industrialized’ social media platforms through a review several major studies of content moderation at scale, including automated systems, machine learning, and fact-checking. The arguable inefficacy and obvious challenges to undertaking these strategies pointed my research toward the role of social psychological dynamics at play in some of the inputs upon which the platforms may rely. Key sources of data collected and used by the systems described in these studies may be subject to the influences of group membership that I discuss in subsequent Parts Three and Four—but which may not be readily apparent to the algorithms or their programmers.

The implementation strategies for content moderation policy at the major social media companies reflect several key circumstances. York and McSherry describe how social media companies have developed those strategies in response to pressures from governments and certain sectors of the public—not least “in the hope of avoiding legal

⁴⁵⁸ Meta, ‘Information on prohibited ads related to voting and ads about social issues, elections or politics’ (*Meta Business Help Centre*) <<https://www.facebook.com/business/help/253606115684173>> accessed 28 August 2022

⁴⁵⁹ Sara Fischer, ‘Technical glitch in Facebook’s ad tools creates political firestorm’ (*Axios*, 30 October 2020) <<https://www.axios.com/2020/10/30/technical-glitch-facebooks-ad-tool-political-election>> accessed 28 July 2022

⁴⁶⁰ Andrew Griffin, ‘Facebook puts groups on “probation” as it tries to stop dangerous content in wake of election results’ (*Independent*, 09 November 2020) <<https://www.independent.co.uk/tech/facebook-groups-us-election-probation-b1719839.html>> accessed 28 July 2022

⁴⁶¹ Facebook, ‘Taking Action Against People Who Repeatedly Share Misinformation’ (26 May 2021) <<https://about.fb.com/news/2021/05/taking-action-against-people-who-repeatedly-share-misinformation/>> accessed 28 July 2022

mandates.”⁴⁶² The largest online intermediaries have shifted away from the early models of content moderation. In the mid-2000s, Google still relied on small moderation teams and in-house legal counsel to make decisions about permissible online expression.⁴⁶³ But the growing scale of the moderation task has led companies to pursue first outsourcing and then automation.⁴⁶⁴ In pursuit of moderation at scale, the companies have instituted systems that are increasingly more secretive, “in which false positives may never see the light of day.”⁴⁶⁵ Yet for the social media companies, the bottom line economics of automating moderation are ineluctable.⁴⁶⁶

The very large datasets generated by social media platforms have long been highly valued by data analysts—a resource described as the “wisdom of crowds” and produced as the users of social networks continuously provide additional data to the system and make connections between existing data and other users.⁴⁶⁷ Information systems specialists claim that “the user-as-information producer model [of social media] provides researchers and news organizations alike with a means of instrumenting and observing a representative sample of the population in real time,” drawing on data from platforms like Twitter to analyse political campaigns, for instance.⁴⁶⁸ The systematized means of news diffusion on a platform like Twitter (i.e., ‘re-tweets’) allowed researchers like Ratkiewicz *et al* to look for common characteristics in sharing metadata to compare how ‘grassroots’ or spontaneous political discussion propagates on the network with the spread of ‘astroturf’ or orchestrated campaigns.⁴⁶⁹ In other words, the traffic patterns of a ‘viral’ post on social media may say a lot about whether it emerged organically or was engineered in a laboratory. However, such studies of information propagation face a “serious obstacle” where the underlying structure of the social network is unknown—and even where relations between users are explicit

⁴⁶² Jillian C York and Corynne McSherry, ‘Content Moderation is Broken: Let Us Count the Ways’ (*Electronic Frontier Foundation*, 29 April 2019) <<https://www.eff.org/deeplinks/2019/04/content-moderation-broken-let-us-count-ways>> accessed 28 July 2022

⁴⁶³ *ibid*

⁴⁶⁴ *ibid*; this development comports with broader economic trend data showing that over the past four decades, automation accounts for a far greater degree of the displacement of routinized tasks than offshoring; see generally Daron Acemoglu and Pascual Restrepo, ‘Tasks, Automation, and the Rise in US Wage Inequality’ (*National Bureau of Economic Research*, June 2021) NBER Working Paper 28920 <<https://www.nber.org/papers/w28920>> accessed 01 December 2022

⁴⁶⁵ York & McSherry (n 462)

⁴⁶⁶ *ibid*; see also Zachary Mack, ‘Why AI can’t fix content moderation’ (*The Verge*, 02 July 2019) <<https://www.theverge.com/2019/7/2/20679102/content-moderation-ai-social-media-behind-the-screen-sarah-t-roberts-vergecast>> accessed 01 December 2022 (discussing with Professor Sarah Roberts the trend away from human content monitoring to scalable computation systems)

⁴⁶⁷ Wei Tan and others, ‘Social-Network-Sourced Big Data Analytics’ (September/October 2013) *IEEE Internet Computing* 62, 62-64

⁴⁶⁸ J Ratkiewicz and others, ‘Detecting and Tracking Political Abuse in Social Media’ (2011) *Proceedings of the Fifth International Association for the Advancement of Artificial Intelligence Conference on Weblogs and Social Media* 297

⁴⁶⁹ *ibid* 297-99

(such as who follows who on Twitter), the strength or importance of those relationships may remain unknown and variable.⁴⁷⁰

Other ‘big data’ approaches to automatic classification focus on the user-generated content itself rather than the topology of the user network. Qazvinian and others manually annotated a dataset of 10,000 tweets related to five different controversial topics for their study into the identification of rumours on social media.⁴⁷¹ The annotations generated extra data about the content: whether the tweet was “rumorous” or not (based on external criteria from an “urban legends” reference site) and whether the language in the tweet endorsed or refuted the rumour (e.g., the presence of negation words like ‘not’ and other features like capitalization and hashtags).⁴⁷² The researchers then used the annotation data to create statistical models designed to predict the likelihood a given tweet was a rumour the tweeter believed based on its corresponding lexical features.⁴⁷³ Such approaches notably face challenges related to unconventional grammar and creative use of language online as well as sarcasm and humour.⁴⁷⁴ Reliance on fixed external references in the training data also introduces a risk of models growing stale when generalized to new rumours and trending topics.⁴⁷⁵ By contrast, the unusual network features of coordinated deception campaigns (if detected early enough) may be distinguishable from organic discussion, even where sentiment analysis or user annotations are less discriminative.⁴⁷⁶

Moderation at scale continues to be the focus in recent social media research. Projects such as the collaboration between Google’s think-tank spinoff Jigsaw LLC and the RAND Corporation have aimed at improving machine learning models for detecting conspiracy theories.⁴⁷⁷ While humans bring valuable contextual knowledge to the ‘conspiracy text’ identification task, these models mechanize the process for deployment at the scale and speed of content generation on global platforms like Google’s.⁴⁷⁸ The Jigsaw-RAND project uses ‘text-mining’ approaches that search for “hedging language” (words and phrases like ‘I think’ and ‘maybe’) and markers of “epistemic certainty” (phrases like ‘we know’ or ‘there is’) that suggest the rhetorical stance of the content producer.⁴⁷⁹ Not unlike

⁴⁷⁰ Ratkiewicz and others (n 468) 298

⁴⁷¹ Vahed Qazvinian and others, ‘Rumor has it: Identifying Misinformation in Microblogs’ (2011) Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing 1589

⁴⁷² *ibid* 1591-92

⁴⁷³ *ibid* 1592

⁴⁷⁴ Qazvinian and others (n 471) 1590

⁴⁷⁵ *ibid* 1596-98

⁴⁷⁶ Ratkiewicz and others (n 468) 302-03

⁴⁷⁷ William Marcellino and others, ‘Detecting Conspiracy Theories on Social Media’ (*RAND Corporation*, 2021) <<https://www.rand.org/t/RR676-1>> accessed 28 July 2022

⁴⁷⁸ *ibid* 2

⁴⁷⁹ *ibid* 3-4

the earlier study by Qazvinian *et al*,⁴⁸⁰ the Jigsaw-RAND study combined human-trained machine classification of semantic content with this statistical model of rhetorical features.⁴⁸¹ The different ‘weights’ they assigned to the rhetorical stance identifiers allowed the RAND researchers to better predict whether particular online “conspiracy talk” tended to promote a conspiracy theory or simply discuss or even oppose it.⁴⁸² While noting their hybrid model would still be susceptible to errors arising from “longitudinal drift” in conspiracy theory content over time, the RAND researchers expect their analysis of rhetorical stance features will prove a durable improvement in a crucial regard—reducing false positives.⁴⁸³ “This is particularly important because an over-zealous system that wrongly flags innocuous or even beneficial talk could result in more harm than good,” they contend, especially “where factual discourse on a topic includes phrases, keywords, or sentiments that overlap with conspiracy.”⁴⁸⁴

Research by Horne, Norregaard, and Adali looked specifically at how well automated fake news detection performs over time.⁴⁸⁵ Horne’s group studied ‘veracity classifiers’—computer tools that aim to perform automated detection and approximations of news veracity based on signals including “article content, source meta-data, external fact-checkers, and crowd sharing behaviour.”⁴⁸⁶ Such methods presume that fake news sources produce “content that it easily distinguished from well-established mainstream news sources” through indicators like simpler language or more negative tone.⁴⁸⁷ In their review of supervised machine learning models,⁴⁸⁸ Horne’s team identified inherent difficulties in the approach—major news events do not necessarily share features with past events contained in the training data (the issue of content “drift”), and unreliable news sources may show fewer stylistic differences with mainstream sources over time.⁴⁸⁹ While these researchers contemplated fake news becoming more sophisticated,⁴⁹⁰ mainstream news has increasingly

⁴⁸⁰ Qazvinian and others (n 471) discussed *supra*

⁴⁸¹ Marcellino and others (n 477) 3-4, 27-28 (also using a collection of Tweets as their dataset); *ibid* 14-16, 29-32

⁴⁸² *ibid* 4-5, 33-39

⁴⁸³ *ibid* 38-39

⁴⁸⁴ *ibid* 39

⁴⁸⁵ Benjamin D Horne, Jeppe Norregaard, and Sibel Adali, ‘Robust Fake News Detection Over Time and Attack’ (2019) 11(1) Association for Computing Machinery Transactions on Intelligent Systems and Technology, article 7

⁴⁸⁶ *ibid* 2

⁴⁸⁷ *ibid*

⁴⁸⁸ Like those studied by Qazvinian and others, Marcellino and others, above: training data (set of news articles pre-screened by fact-checkers) is provided to the computer model, which classifies news articles based on content labels; the model is then tested in a ‘lab setting’ to see how accurately it classifies additional news stories

⁴⁸⁹ Horne, Norregaard, and Adali (n 485) 2-7

⁴⁹⁰ *ibid* 2, 9

become less sophisticated in order to grab attention and chase clicks,⁴⁹¹ further shrinking the lexical and stylistic differences between the two. Horne's group found that the performance of supervised machine learning models degrades over time, especially as the news cycle shifts.⁴⁹² These models essentially require frequent 'retraining' to keep up with novel information.⁴⁹³ Interestingly, blocking unreliable news sources from algorithmic consideration made subsequent fake news detection dramatically less accurate,⁴⁹⁴ perhaps a sort of computer science verification of the old adage 'better the devil you know.'

Government policy (including the looming prospect of regulation) pressures social media companies towards increasing levels of moderation as they scale up their user bases.⁴⁹⁵ In the context of news content policy, a key aspect of that pressurization is time sensitivity. Existing content regulation like Germany's *NetzDG* provisions set strict requirements for timely responses by platforms when users post proscribed content.⁴⁹⁶ Elections-related news content has its own inherent time characteristics: the novelty to the news cycle of news stories about candidates, and the window of relevance of such stories during the period leading up to the election itself.

The variety and multiplicity of news sources that can pursue audiences through social media compounds the aspect of time sensitivity, for both platform engineers and government policymakers. More sources appear to be able to say more things about election candidates and be heard by more people more quickly, as discussed in Part One.⁴⁹⁷ Accordingly, a political and technical focus on authoritative sources has emerged as a means for achieving timely responsiveness. Whether as 'recognized news publishers' under the Online Safety Bill or the EC's 'high quality journalism', the State of Florida's 'journalistic enterprises' or Canada's 'legitimate sources of information', governments evidently view source legitimation as an important tool for managing the flow of information online.⁴⁹⁸ Social media platforms look to authoritative sources (through institutions like third-party fact-checkers) as a crucial data resource in training their content moderation algorithms.

However, both research and practical experiences have revealed important considerations about the inefficacy or unreliability of a narrow focus on authoritative sources. In their study of the effects of source ratings on fake news stories presented as

⁴⁹¹ See Part I, 'Social Media News Use' and Part I, 'Fake News Studies'

⁴⁹² Horne, Norregaard, and Adali (n 485) 11-12

⁴⁹³ *ibid* 12-14, 19-20

⁴⁹⁴ *ibid* 18-19

⁴⁹⁵ Caplan (n 376), York and McSherry (n 462)

⁴⁹⁶ Google, 'Removals under the Network Enforcement Law' (n 392) ('obviously illegal' content must be taken down within 24 hours of a complaint)

⁴⁹⁷ Part I, 'Social Media News Use'

⁴⁹⁸ Discussed in further detail in Part V, 'Public Policy and the Public Sphere'

Facebook posts,⁴⁹⁹ Kim and Dennis found that while news source reputation had a significant relationship with the believability of a story, that reputation did not significantly relate to whether users liked or shared the post.⁵⁰⁰ Instead, Kim and Dennis found that users are significantly “more likely to read, like, and share articles that match their point of view,”⁵⁰¹ and that making news sources more salient “nudges readers to be more skeptical of *all* articles, regardless of their source.”⁵⁰² Discoveries like these speak to a deeper role for social media user psychology in how platform technical provisions actually operate, which I examine further in the rest of this Thesis.

Third party fact-checkers often serve as the putative basis for determinations of news source reliability, both in studies⁵⁰³ like Horne, Norregaard, Adali’s and in the technical solutions deployed by social media platforms like Facebook.⁵⁰⁴ Fact-check companies like NewsGuard use “trained journalists to rate thousands of news and information websites” according to various criteria to label sites as “unreliable.”⁵⁰⁵ In the company’s 2021 year-end report, NewsGuard provided a list of its “Ten Top ‘Trustworthy and Trending’ Sites”—including American newspaper USA TODAY at number five.⁵⁰⁶ Yet in 2022, USA TODAY announced that an internal audit had uncovered twenty-three articles on its website that were literally fake news—

“The audit revealed some individuals quoted were not affiliated with the organizations claimed and appeared to be fabricated. The existence of other individuals quoted could not be independently verified. In addition, some stories included quotes that should have been credited to others.”⁵⁰⁷

The Washington Post, ranked even higher by NewsGuard,⁵⁰⁸ ran a story in March 2021 that falsely attributed statements to U.S. President Donald Trump about “fraud” in the 2020 election process in the State of Georgia “based on information provided by a source” which

⁴⁹⁹ Antino Kim and Alan R Dennis, ‘Says Who? The Effects of Presentation Format and Source Rating on Fake News in Social Media’ (2019) 43(3) *Management Information Systems Quarterly* 1025

⁵⁰⁰ *ibid* 1032-33

⁵⁰¹ *ibid* 1032

⁵⁰² *ibid* 1034 (emphasis in original)

⁵⁰³ Horne, Norregaard, and Adali (n 485) 5-6

⁵⁰⁴ Atlantic Council (n 454)

⁵⁰⁵ Sara Fischer, “‘Unreliable’ news sources got more traction in 2020’ (*Axios*, 22 December 2020) <<https://www.axios.com/2020/12/22/unreliable-news-sources-social-media-engagement>> accessed 28 July 2022

⁵⁰⁶ Matt Skibinski, ‘Special Report: The U.S. Best and Worst of 2021’ (*NewsGuard*, 29 December 2021) <<https://www.newsguardtech.com/special-reports/special-report-the-u-s-best-and-worst-of-2021/>> accessed 01 December 2022

⁵⁰⁷ USA TODAY, ‘USA TODAY removed 23 stories from website, other platforms following audit of reporter’s work’ (*USA TODAY*, 16 June 2022) <<https://eu.usatoday.com/story/news/2022/06/16/usa-today-audit-reporter/7647731001/>> accessed 01 December 2022

⁵⁰⁸ Skibinski (n 506)

the Post relayed anonymously.⁵⁰⁹ The Post provided a correction to the story months later after an audio recording proved the supposed statement was fabricated.⁵¹⁰

Given that news source may not be foremost in the minds of social media users sharing news stories with one another, and that even highly-rated sources can commit egregious journalistic errors, fact-checking appears to play an unusually prominent role in discussions of content moderation. This role is materially constituted in the EC's Code of Practice on Disinformation, for example, as well as the various partnerships between the platforms and third parties, as discussed previously. Despite early expectations⁵¹¹ that the Internet would empower 'citizen journalists' to contribute to an enriched public sphere, and seemingly against the express preferences of news consumers seeking alternative points of view online, the digital extensions of large media conglomerates dominate the rankings of fact-checkers like NewsGuard.⁵¹² Nearly all of NewsGuard's top ten "trustworthy" news sites are household names to consumers of newspapers or television news. Few if any of its ten "most influential misinformers" are widely read or have any substantial offline footprint.

The role of third-party fact-checkers thus appears to be to reproduce existing material relationships rather than getting at the truth of a news event. Perhaps the obvious way to get at the truth of a news event would be to have *more* disinterested journalists able to investigate it. Third-party fact-checkers instead participate in the kind of credibility management game that Greenwald describes as common to institutionalized players in the news business, helping entrenched interests to close ranks and exclude outside competitors.⁵¹³ This role is part of why third-party fact-checkers are valuable not only to legacy media but especially to governments and national security services. Studies of the media by Hermann and Chomsky have described how "the large bureaucracies of the powerful *subsidize* the mass media, and gain special access by their contribution to reducing the media's costs of acquiring the raw materials of, and producing, news."⁵¹⁴ Fact-checkers may similarly raise the cost of establishing credibility for independent and critical news

⁵⁰⁹ Amy Gardner, 'Trump pressured a Georgia elections investigator in a separate call legal experts say could amount to obstruction' (*Washington Post*, 11 March 2021) <https://www.washingtonpost.com/politics/trump-call-georgia-investigator/2021/01/09/7a55c7fa-51cf-11eb-83e3-322644d82356_story.html> accessed 28 July 2022

⁵¹⁰ *ibid* ('Correction' notice)

⁵¹¹ Part I, 'Social Media News Use'

⁵¹² Skibinski (n 506)

⁵¹³ Glenn Greenwald, 'How Do Big Media Outlets So Often "Independently Confirm" Each Other's Falsehoods?' (*Substack*, 16 March 2021) <<https://greenwald.substack.com/p/how-do-big-media-outlets-so-often>> accessed 01 December 2022

⁵¹⁴ Edward S Hermann and Noam Chomsky, *Manufacturing Consent* (Pantheon Books, updated ed. 2002)

sources and lower it for sources that buttress opinions which are “functional for elite interests.”⁵¹⁵

These relationships are replicated by researchers, governments, and corporate policymakers who rely on third-party fact-checking assessments to frame the appropriateness of online news discourses. The online media reproduction of offline media materiality—domination by a small number of outlets and dependence on government bonhomie—consequently reproduces online the dissatisfaction and distrust that so many people around the world report regarding offline news outlets as discussed in the previous Part.⁵¹⁶ Boyd has discussed similar consequences arising from media literacy campaigns against misinformation: the potential for backfire where media literacy ‘solutions’ are merely cover for the worldview of those in authority, rather than a means for interrogating how different groups think differently from different perspectives.⁵¹⁷ Continuing or growing public dissatisfaction with the news discourses available on social media in turn will further pressure social media companies to develop more responsive tools for identifying and managing content on their platforms.

Ultimately, for the platforms, the broadest, timeliest, and most accessible data on user-generated content comes from the users themselves. Supervised machine learning methods cannot distinguish between “a well-written lie and truth,” such as when a fake news source produces content that “reads very much like one produced by a mainstream source.”⁵¹⁸ Conversely, a poorly-written truth may not be readily distinguished from falsity by a machine. The truth can be written in simple language and negative tone—those are not intrinsically dispositive features but *are* the kinds of features acted upon by such models. Fact-checking, regardless of the interests it may serve, necessarily lags behind the emergence and dissemination of novel news narratives. The dialogue between users and the platforms, through the processes of content flagging and reporting,⁵¹⁹ will remain the most immediate and available resource for social media companies in meeting time and scale constraints.

Yet even with such a robust resource, social media content moderation at scale faces persistent obstacles. York and McSherry point out several ways that content moderation is inherently inconsistent.⁵²⁰ For one, the more visible a user’s content is, the more likely it is

⁵¹⁵ Hermann and Chomsky (n 514) 23 (emphasis in original)

⁵¹⁶ Part I, ‘Social Media News Use’

⁵¹⁷ danah boyd, ‘You Think You Want Media Literacy... Do You?’ (*Data & Society: Points*, 9 March 2018) <<https://points.datasociety.net/you-think-you-want-media-literacy-do-you-7cad6af18ec2>> accessed 01 December 2022

⁵¹⁸ Horne, Norregaard, and Adali (n 485) 20

⁵¹⁹ Crawford and Gillespie (n 319) 412

⁵²⁰ York and McSherry (n 462)

to be policed.⁵²¹ Likewise, moderation processes typically cannot discriminate *why* a user has shared a particular piece of content, such as for ‘calling out’ offensive content.⁵²² Bernal has pointed out instances where content policies aimed at fighting abuse by right-wing online ‘trolls’ have resulted in suspensions and bans of leftist activists just as readily.⁵²³ Furthermore, adequate appeals processes concerning automated systems may remain difficult to achieve, especially due to the challenge of making such systems transparent.⁵²⁴ In March 2021, Twitter users experienced a wave of bans from the service for tweeting the word “Memphis,” possibly as a result of the word’s inclusion in Twitter’s automated systems for spotting problematic content⁵²⁵—“due to a bug” according to Twitter.⁵²⁶

This Part has provided a basis for considering which elements of the social media system resemble those of past media systems, particularly where they touch on broader civic life as in news production, consumption, and distribution. Rather than radically displacing the traditional news industry, the largest social media platforms feature the intensified centralization of ‘editorial’ decision making about news content—yet embody a new ethos that prizes broad networking over vertical hierarchy. In the regulatory models currently developing in the U.S., U.K., and more broadly, policymakers typically look to the major social media companies to do the actual work of evaluating user-generated content. The pressure for results applied to the platforms is especially acute in the consequential and time-sensitive context of elections. Accordingly, I have also examined some of the internal policies and techniques developed by YouTube, Twitter, and Facebook for governing the kinds of elections news content they will distribute or sanction. At the vast scale of their operations, this governance is achieved primarily through automated content moderation and reliance on the huge amounts of data provided to the platforms by their users. These capabilities bear directly on the kinds of implementations which both policymakers and the platforms view as feasible for managing ‘fake news’ content.

Studies like that by Kim and Dennis,⁵²⁷ however, point to a substantial source of error that may persist despite content moderation systems working as intended. Before any data reaches the weighted decision trees and promotion algorithms of the platforms, users make

⁵²¹ York and McSherry (n 462)

⁵²² *ibid*

⁵²³ Paul Bernal, ‘The push to ban Nazis on Twitter can just as easily affect the left – as the suspension of a Corbyn supporter proves’ (*Independent*, 05 March 2019) <<https://www.independent.co.uk/voices/twitter-ban-nazi-suspend-rachael-swindon-labour-corbyn-a8808156.html>> accessed 01 December 2022

⁵²⁴ York and McSherry (n 462)

⁵²⁵ Danielle Zoellner, ‘Twitter users flag bug where accounts are banned if they tweet “Memphis”’ (*Independent*, 14 March 2021) <<https://www.independent.co.uk/tech/twitter-accounts-ban-memphis-b1817064.html>> accessed 01 December 2022

⁵²⁶ Twitter Support (*twitter.com*, 14 March 2021)

<<https://twitter.com/twittersupport/status/1371235587718582274>> accessed 01 December 2022

⁵²⁷ Kim and Dennis (n 499) 1034

decisions about the content they see. I have surmised that this may play a substantial role in whether solving the fake news problem can be achieved without undue interference on genuine but politically charged news content on social media. Recognizing further that the systemic ‘sticking point’ of mass-scale automated moderation may reflect an important human, social psychological component, I determined to conduct further research that could highlight that component—and in turn to evaluate whether policy proposals sufficiently account for any such aspect of social media operation.

In order to better understand how social media users make judgments about the authenticity of political news, I decided to create an experiment that would test people’s decision-making about the kinds of stories they might see online during the run-up to an election. The data generated by my experiment aims to approximate the kinds of data social media companies rely on to label certain user-generated content as problematic, especially in the automated processes the companies use which I explored in this Part. My experimental research also takes the approach of avoiding the morass of arbitrarily defining ‘fake news’ and reliance on observations that occur after the social media platforms have applied their processes of moderation and promotion. Based on the social psychological principles I discuss in Part Three, I also made key inferences about how people’s political affiliations would relate to their ‘fake news’ judgments. The design and results of my experiment to test those inferences are covered in Part Four.

Key Takeaways

- Answers to problems of disruptive influence arising in past media were adopted and shed as circumstances (including structural factors and social relations) changed
- The social media system features more highly concentrated and automated control over news distribution than past media systems and even other Internet media
- Scrutinizing this aspect of social media more closely will help us understand whether past answers will effectively address current problems of disruptive influence

Part Three: The Social Psychology of Decision Making

Part Three of this Thesis concerns the role of social psychology in making decisions about news content on social media. My goal here is to establish social psychology as a potentially significant constraint in efforts to address fake news through the automated moderation systems that are integral to content policy implementation at the major social media platforms. The first section of this Part introduces psychology in relation to the legal system. The next section reviews some of the past research that established key concepts in social psychology, including social identification and various accounts of motivated decision making. The third section of this Part covers specific research into the social psychology of media use. The research reviewed in this section provides insights into the human factors that may have frustrated some of the media studies and computer science research discussed in Part Two. In the final section of this Part, I discuss how these insights have informed my research into how political group identification relates to the kind and quality of information used in content moderation. In Part Four, I explain my research design and analyse the results of the experiment I conducted to test my research question. These Parts provide a more nuanced understanding about the most distinctive technical aspect of the major social media platforms, and in turn open more robust ways of assessing the public policy proposals to address fake news on these platforms, as discussed in Part Five.

Introduction to Law and Psychology

Courts have long attempted to understand the consequences of human thinking and emotion on the evaluation of evidence and the exercise of judgment in legal proceedings, promulgating varied perspectives on the efficacy of jury trials⁵²⁸ and creating rules of evidence to guard against unfair prejudice.⁵²⁹ Under the U.S. Federal Rules of Evidence, for example, evidence of “a person’s character or character trait is not admissible to prove that on a particular occasion the person acted in accordance with the character or trait.”⁵³⁰ Not without exceptions, the Rule attempts to balance the relevance of character evidence with its tendency to prejudice a trier of fact’s evaluation of the case at hand. While the Rule reflects the long accumulation of judicial experience, psychologists have brought scientific

⁵²⁸ Lady Justice Hallett, ‘Trial by Jury – Past and Present’ (35th Blackstone Lecture, 22 May 2017) <<https://www.judiciary.uk/wp-content/uploads/2017/05/hallett-lj-blackstone-lecture-20170522-1.pdf>> accessed 31 May 2020.

⁵²⁹ FED. R. EVID. 403

⁵³⁰ FED. R. EVID. 404(a)

approaches to bear on similar questions of how we assess evidence in light of our preconceptions and circumstances.⁵³¹

In the 1970s, researchers such as Loftus began examining how eyewitnesses' memories can change, how line-up procedures unintentionally influence suspect identification, and other questions about the reliability of witness testimony.⁵³² This line of investigation has helped uncover the "myriad ways in which a person might come to believe and report a misinformation memory. . . . Whatever process is going on, the end result can be a highly confident witness testifying in a persuasive manner at trial about a detail that is completely false."⁵³³ Courts have to reckon with the challenging fact that a witness may not be a reliable narrator of even his own involvement in an incident, much less another person's.⁵³⁴ Although expert testimony from psychologists is not always available (or necessarily appreciated),⁵³⁵ psychologists support the forensic function of the legal system by helping judges and juries understand the role and likelihood of psychological effects.⁵³⁶

By the 1980s, the relationship between psychology and the legal system had begun to generate significant research interest beyond the forensic application of psychological expertise.⁵³⁷ Psychologists such as Haney envisioned a broader application of psychological principles to the legal system with an eye towards legal change or reform—psychology *and* the law or a psychology *of* the law, rather than merely within it.⁵³⁸ Subsequent research in this focus has interrogated the "extralegal" assumptions about behaviour that play a role in legal processes and outcomes.⁵³⁹ For example, Lindholm and Christianson studied how eyewitness accounts of a violent crime appeared "systematically biased as a function of the group membership of the perpetrator," finding that "participants rated a perpetrator who stabbed a victim . . . as more culpable if he was ethnically dissimilar" to themselves than otherwise.⁵⁴⁰ As lawyers, judges, and legal scholars have become more aware of these issues,

⁵³¹ The tension between the cut-and-dried facts of a case and the myriad beliefs and prejudices of its participants have likewise featured in dramatic expositions of the law, from Shakespeare's rendition of Henry VIII and Katherine's divorce (*Henry VIII*, II. 4) and Shylock's proceedings against Antonio (*The Merchant of Venice*, IV. 1.) to Sidney Lumet's *12 Angry Men* (United Artists, 1957)

⁵³² Elizabeth F Loftus, 'Eyewitness testimony' (2019) 33 *Applied Cognitive Psychology* 498, 499-500 (reviewing her own work and that of others in the field)

⁵³³ *ibid* 499

⁵³⁴ Cécile Schain and others, 'Looking at the actor's face: Identity cues and attentional focus in false memories of action performance from observation' (2012) 48 *J Experimental Social Psychology* 1201, 1203-04

⁵³⁵ Loftus (n 532) 500-01

⁵³⁶ Ronald Blackburn, 'What is forensic psychology?' (1996) 1 *Legal and Criminological Psychology* 3, 11-12

⁵³⁷ See generally Craig Haney, 'Psychology and Legal Change: On the Limits of a Factual Jurisprudence' (1980) 4(3) *Law and Human Behaviour* 147 (discussing the development of 'law and psychology')

⁵³⁸ *ibid* 150

⁵³⁹ *ibid* 154-55

⁵⁴⁰ Torun Lindholm and Sven-Ake Christianson, 'Intergroup Biases and Eyewitness Testimony' (1998) 136(6) *J Social Psychology* 710, 720; *cf.* Rebecca M Nichols and Elizabeth F Loftus, 'Who is susceptible in three false

they have attempted to anticipate and manage the ways bias, prejudice, and discrimination can impact criminal and civil proceedings.⁵⁴¹ Further efforts to integrate behavioural analysis, law, and economics reimagine laws and regulations as optimizable ‘nudges’ for producing desired behavioural results from particular “cognitive mechanisms.”⁵⁴² This vision carries its own set of normative prescriptions about what constitutes optimal outcomes,⁵⁴³ and as Frerichs demonstrates in her survey of the field, a narrow focus on proximate explanations of behaviour tends to obscure or ignore social factors like institutional or political power.⁵⁴⁴

Another avenue of psychological investigation instead emphasizes the social dimension of human thinking, feeling, and interaction and explicitly recognizes that individual decision making depends on variable social situations.⁵⁴⁵ Frerichs notes that in “applying this to the law, the specific social, contextual, and situational factors shaping individual decision-making would gain more analytical weight compared to (over)generalised accounts of cognitive biases, dispositions or mindsets as individual properties.”⁵⁴⁶ The value of psychological research in this sense can come from showing us where to look and suggesting what questions to ask about the social interactions that give rise to legal consequences. My approach in this Thesis is to emphasize that psychology can tell us much about what’s going on during human interactions, while what the law should say or do about it remains bound together with politics, values, and trade-offs between alternatives. In this and subsequent Parts, I aim to provide a scientific basis for understanding specific behaviours involving political news on social media and to assess how various expressions of the law may operate in light of that understanding.

Principles of Social Psychology

Social psychology has developed as a specific discipline over the course of decades through thought provoking explorations of how social life influences individual thinking, feeling, and behaviour. For the purposes of this Thesis, ‘social psychology’ refers to the study of individual behaviour in social contexts. Taylor and Moghaddam note:

memory tasks?’ (2019) 27 (7) *Memory* 962 (noting that individual differences do not appear to explain the widely observed phenomena of eyewitness memory impairment or false memories)

⁵⁴¹ Jerry Kang and others, ‘Implicit Bias in the Courtroom’ (2012) 59 *UCLA L Rev* 1124

⁵⁴² Sabine Frerichs, ‘Putting behavioural economics in its place: the new realism of law, economics and psychology and its alternatives’ (2021) 72(4) *NILQ* 651, 666-67, 675 (describing the field of behavioural law and economics and its critique of assumptions in standard economic models of law and regulation)

⁵⁴³ *ibid* (the expert ‘nudgers’ know what’s in your “own best interest” or desirable and what is “suboptimal”)

⁵⁴⁴ *ibid* 669-72 (contrasting behavioural economics and law with socio-legal studies and socioeconomics’ attention to social context)

⁵⁴⁵ *ibid* 678-79

⁵⁴⁶ *ibid* 679

“Although the unit of analysis in social psychology is the individual, the context of such behaviour can involve, either implicitly or explicitly, both other individuals and social groups. The unique perspective of social psychology in an intergroup context is that the perceptions, motivations, feelings, and overt actions of individuals are studied to identify how they influence, and are affected by, relations between groups.”⁵⁴⁷

The extent to which individual differences or dispositions give rise to cognition and behaviour has long been the subject of extensive inquiry and philosophical debate.⁵⁴⁸ Early experiments in social psychology identified an array of distinctly *interpersonal* forces that influenced how test subjects evaluated information derived from observations.⁵⁴⁹ The social forces of interest to these researchers included the complexity of the experimental situations, the role of confidence and trust in others, authority, and pressure to conform.⁵⁵⁰ In the 1950s, Asch conducted an experiment which involved generating a disagreement between an individual test subject and a group of confederates⁵⁵¹ concerning a clear matter of fact in the participants’ immediate environment: the visually apparent length of a series of lines.⁵⁵² In the test condition, the confederate group publicly stated a wrong judgment about the length of the lines and then the test subject announced his own length judgment. Asch found that these test subjects showed “marked movement toward the majority” in their assessments, compared to nearly errorless judgments of line length by subjects in the control group (where the confederate group did not announce their ‘wrong’ assessment).⁵⁵³ The impressive result of the experiment, along with others of the 1950s and 1960s, heralded extensive investigation over the subsequent decades to try to understand why people’s senses, judgments, and actions are apparently so influenced by their social context.

In addition to social pressure, social psychologists have extensively researched attribution and stereotyping—two processes with important roles in intergroup relations. Attribution refers to “how we make judgments about other people, ourselves as well as others,” particularly with regard to “making a judgment about the cause or causes of

⁵⁴⁷ Donald M Taylor and Fathali M Moghaddam, *Theories of Intergroup Relations: International Social Psychological Perspectives* (2nd edn, Praeger 1994)

⁵⁴⁸ W S Taylor, ‘Is Truth Individual or Social?’ (August 1935) *J Social Psychology* 348

⁵⁴⁹ Abraham S Luchins and Edith H Luchins, ‘Social influences on judgment of changing evidence’ (1961) 54 *J Social Psychology* 13, 34-35 (tested what happened when subjects overheard someone else’s response to a stimulus—unbeknownst to the subject, pre-instructed by the experimenter to respond in a certain way)

⁵⁵⁰ *ibid* 30-34

⁵⁵¹ ‘Confederates’ are participants directed by the experimenter

⁵⁵² Solomon E Asch, ‘Studies of Independence and Conformity: I. A Minority of One Against a Unanimous Majority’ (1956) 70(9, Whole No. 416) *Psychological Monographs* 1, 1-2

⁵⁵³ *ibid* 9-11

behavior.”⁵⁵⁴ Attributions can be made to internal or external causes, and the tendency to “underestimate the impact of situation (external) and overestimate the importance of the actor’s traits and attitudes (internal)” is called the fundamental attribution error.⁵⁵⁵ When making attributions about ourselves, we tend to do so in self-serving ways, and this tendency has been the subject of much research into bias.⁵⁵⁶

Stereotyping has often been viewed as an undesirable or negative process, taking the form of overgeneralizing about a particular group.⁵⁵⁷ However, psychologists now understand stereotyping results from a basic cognitive process that is neither inferior to other modes of thinking or morally wrong in-and-of itself.⁵⁵⁸ Rather, we rely on stereotypes as a perceptual heuristic—a mental shortcut to manage the complexity of stimuli in our environment about social groups or individuals as members of groups.⁵⁵⁹ This process operates quickly (and largely outside of conscious awareness) to integrate what we know about groups with who we perceive to be a group member.⁵⁶⁰ We even stereotype ourselves: “a special case of stereotyping where members of a group attribute characteristics to their own group.”⁵⁶¹

Stereotypes result automatically as we take in and mentally organize information about social groups or group members; the process is simplified by “ignoring certain dissimilarities and giving priority to, or exaggerating, certain similarities.”⁵⁶² We do this to some extent in the same way that we categorize objects—when we need to find a hammer, we focus on objects with similar characteristics like being the right size and shape and hardness, and so might pick up an appropriate rock as a substitute, but not a tomato, which we would ignore or quickly reject.⁵⁶³ This categorization process “orients and assists our actions,” speeding up our decision-making by “simplifying the environment, a simplification that is essential if we are to be able to process effectively, and act upon, the potentially endless amount of information available to us.”⁵⁶⁴ In the 1960s and 1970s, Tajfel studied how the categorization process tended to result in perceived uniformity within categories

⁵⁵⁴ Taylor and Moghaddam (n 547) 166

⁵⁵⁵ *ibid* 167-68

⁵⁵⁶ *ibid*

⁵⁵⁷ *ibid* 160-61

⁵⁵⁸ *ibid* 161-62

⁵⁵⁹ Donald M Taylor, *The Quest for Identity: From Minority Groups to Generation Xers* (Praeger 2002) 38

⁵⁶⁰ See, e.g., Fernando Gordillo and Lilia Mestas, ‘What we know about people shapes the inferences we make about their personalities’ (2015) 23 *Current Research in Social Psychology* 38; Jayati Sinha and Dhananjay Naykankuppam, ‘The canny social judge: Predicting others’ attitudes from sparse information’ (2014) 53 *J Experimental Social Psychology* 145

⁵⁶¹ Taylor (n 559) 38-39

⁵⁶² Taylor and Moghaddam (n 547) 67

⁵⁶³ *ibid* 67

⁵⁶⁴ *ibid*

and distinctiveness between categories;⁵⁶⁵ in other words, how various hammers seem more alike to other hammers, and hammers in general seem more different from axes. Subsequent research on categorization has looked at how the process shapes our perception of objects and their categories simply by its operation.⁵⁶⁶

But categorizing social stimuli like someone's apparent ethnicity or regional accent also draws in norms and values associated with the groups in play⁵⁶⁷—special characteristics that people have, but hammers or axes do not. Social categorization has nevertheless been found to function in a basically consistent way, helping us quickly make sense of information in our social environment.⁵⁶⁸ Tajfel and others have described the social categorization process as fundamental to intergroup relations—that indeed “there can be no intergroup behaviour without the relevant aspects of the social environment having been categorized in terms of whatever may be the pertinent social criteria for the lines of division of people into ‘us’ and ‘them’, into groups and outgroups.”⁵⁶⁹

The similar tendency towards distinctive categories among non-social and social stimuli led Tajfel to develop a broader concept called social identity theory to explain what animates the thinking and behaviour of people in relation to groups.⁵⁷⁰ In his experiments, Tajfel had recognized that when social categorization involved the social group of the person doing the categorizing, his participants tended to act in favourable ways towards the ‘ingroup’ with which they identified.⁵⁷¹ Tajfel and others found that even between artificial, arbitrarily classified groups, ingroup preferences emerged without regard to any “‘objective’ conflict of interests or in terms of deep-seated motives” the ingroup bias might serve.⁵⁷² A key insight of this finding is that intergroup behaviour cannot be fully understood in terms of conflict or competition alone; rather, traditional accounts of discrimination, for instance,

⁵⁶⁵ Taylor and Moghaddam (n 547) 67-68

⁵⁶⁶ Sarah Queller, Terry Schell, and Winter Mason, ‘A Novel View of Between-Categories Contrast and Within-Category Assimilation’ (2006) 91 (3) *J Personality and Social Psychology* 406 (investigating how this process relies on feedback loops: what we observe about a thing tells us what the category is like, which tells us what to observe about the thing)

⁵⁶⁷ Taylor and Moghaddam (n 547) 68-69

⁵⁶⁸ *ibid*

⁵⁶⁹ Henri Tajfel and others, ‘Social Categorization and Intergroup Behaviour’ (1971) 1(2) *Eur J Social Psychology* 149, 151; see also Taylor and Moghaddam (n 547) 6, defining intergroup relations as “any aspect of human interaction that involves individuals perceiving themselves as members of a social category, or being perceived by others as belonging to a social category.”

⁵⁷⁰ See generally Henri Tajfel, ‘Social Psychology of Intergroup Relations’ (1982) 33 *Annual Review of Psychology* 1, 24

⁵⁷¹ Taylor and Moghaddam (n 547) 70-73

⁵⁷² Tajfel and others (n 569) 176

should be complemented with an awareness of how social categorization itself can drive behaviour.⁵⁷³

Social Identity Theory

The central proposition of social identity theory is that “people desire to belong to groups that enjoy distinct and positive identities.”⁵⁷⁴ Social identity theory postulates this is so because the same tendency that pushes us towards more consistent, distinct categories for the stimuli in our environment also draws in information about the groups to which we belong. Because of this, we tend to notice the things that members of our group have in common and the things that make our group stand out from other groups. And because our associations with social stimuli also draw in assessments of group norms and values, our mental picture of a particular group can hold quite a lot of information. Just from categorizing a person as belonging to ‘Group A’, we might have a host of expectations of what Group A people are like and how they behave, and we rely on those in deciding whether to approach this Group A person and how to interact. Social categorization likewise directly interacts with how we see ourselves—the ‘social identity’ each of us holds, made up of our knowledge of belonging to a particular group and the positive or negative values we attach to membership in that group.⁵⁷⁵

This interaction between social identity and social categorization helps explain why different aspects of our identities feel more relevant in specific contexts. When group membership becomes salient—in other words, especially apparent or important—the information that individuals consider relevant changes: When thinking about our group, we tend to notice certain social stimuli that we might ignore in a context where our group membership is not salient.⁵⁷⁶ Two hypothetical Aviva employees Tom and Paul might see themselves quite similarly as co-workers when chatting with a Lloyds employee over lunch in the Norwich city centre, but relate rather differently to one another on match days as supporters of rival football clubs. Over lunch, Tom and Paul might compare the same corporate logo on their business cards with the Lloyds employee’s, converse in common terms of shared workplace norms or practices, and will generally see each other as members of the same group, if it crosses their minds at all. But on match days, Tom wearing green and yellow stands out vividly to Paul, who’s wearing blue and white, and vice versa; and each

⁵⁷³ Tajfel (n 570) 24

⁵⁷⁴ Taylor and Moghaddam (n 547) 62

⁵⁷⁵ *ibid* 78

⁵⁷⁶ Angela T Maitner and others, ‘Identity salience moderates processing of group-relevant information’ (2010) 46 *J Experimental Social Psychology* 441, 441-442, 444 (a news item about changes to the university examination process will stand out more when readers are thinking of themselves as students rather than thinking about themselves just as Americans or Canadians, for example)

might think the other seems a lot more excitable and aggressive than either previously assumed.⁵⁷⁷ These visual and auditory social stimuli engage the social categorization process, and in turn Tom and Paul relate to one another through different group identities on match days than at their business lunch.

Why is identity in general and social identity in particular such a powerful human drive? According to Taylor, identity serves as “our own personal blueprint for action; it guides our thoughts, feelings, and actions in a coherent manner.”⁵⁷⁸ That ‘blueprint’ makes possible our interactions with the world around us by giving us a descriptive account of ‘who I am’ that we can relate to everything else.⁵⁷⁹ Taylor further proposes that personal identity and self-esteem must follow on from collective identity: “Personal identity involves individuals comparing themselves with members of their own group in order to determine what characteristics make them unique. Similarly, personal self-esteem requires that the individual make evaluative comparisons with his or her own group.”⁵⁸⁰ Thus, having a clearly defined social identity would play an essential role in supporting individual development.⁵⁸¹ That appears to be supported by research finding that identifying with our ingroups can enhance our sense of well-being and decrease negative mood in ways that are distinct from the benefits we gain from interpersonal relationships generally.⁵⁸²

Furthermore, our experience of social situations as a group member can engage our emotions *as a group member* rather than just as an individual. Intergroup emotions theory (“IET”) looks at the role of emotions in intergroup relations and behaviour, examining emotional states that arise from “belonging to, and by deriving identity from, one social group rather than another.”⁵⁸³ IET studies have shown that people report feeling different emotions when they think of themselves as group members rather than as individuals, that their responses as group members are shared with one another, and that the members who identify more strongly with their group experience these group emotions relatively more intensely.⁵⁸⁴ Feelings, particularly ones like anger, can influence behaviour, and group

⁵⁷⁷ See, e.g., Yi Jenny Xiao and Jay J Van Bavel, ‘Sudden shifts in social identity swiftly shape implicit evaluation’ (2019) 83 J Experimental Social Psychology 55 (studying how our assumptions about others shift when our context for interaction shifts: biases between arbitrarily assigned groups seem to turn off when groups are reframed as cooperative and even reverse when group membership is switched)

⁵⁷⁸ Taylor (n 570) 35

⁵⁷⁹ *ibid* 36

⁵⁸⁰ *ibid* 40

⁵⁸¹ *ibid* 40-41

⁵⁸² A Kyprianides, M J Easterbrook, and R Brown, ‘Group identities benefit well-being by satisfying needs’ (2019) 84 J Experimental Social Psychology 103836

⁵⁸³ Diane M Mackie, Eliot R Smith, and Devin G Ray, ‘Intergroup Emotions and Intergroup Relations’ (2008) 2/5 Social and Personality Psychology Compass 1866, 1867

⁵⁸⁴ *ibid* 1869

emotions (because of their social source or target) tend to motivate intergroup behaviour.⁵⁸⁵ Rothschild and others examined how scapegoating behaviours may reflect group members' "motivated displacement of blame . . . to reduce elevated feelings of collective guilt for ingroup harm-doing."⁵⁸⁶ Such group emotions only have their motivating force when a person identifies him or herself as part of the group to which they pertain.

Social identity theory also holds that the need for a positive and distinct social identity drives the comparisons we make between groups.⁵⁸⁷ These comparisons reflect the evaluative side of social categorization (whereas identity itself could be thought of as more category-'descriptive') and include our assessments of group status and social position.⁵⁸⁸ Such evaluations include the inferences we make about "interests, motives, intuitions, actions, and attributes of groups and of individuals" in relation to their observed social conduct.⁵⁸⁹ Someone's perceived political orientation, for instance, can shape our beliefs about how willing that person is to listen to new ideas or change their opinions.⁵⁹⁰ When we make attributions about behaviour at the group level, we can similarly tend towards those errors we make at the individual level.⁵⁹¹ Taylor and Moghaddam explain that

Like the self-serving bias, the error in this case operates in such a way that when members of an in-group make attributions about the socially desirable behavior of their own group, they focus on the positive traits (internal), but when they attribute the same desirable behavior to an out-group, an external cause is the focus. Conversely, for socially undesirable behavior, external causes are the focus for the in-group. However, when the undesirable behavior is performed by an out-group, the focus is an internal factor, such as the negative traits of the out-group.⁵⁹²

In an illustrative and thought-provoking example of social identity and social comparison in action, Tarrant and others found that British and American study participants each rated

⁵⁸⁵ Mackie, Smith, and Ray (n 583) 1874; see also Daniel B Shank and Alexander Burns, 'Comparing Groups' Affective Sentiments to Group Perceptions' (2018) 26 *Current Research in Social Psychology* 55

⁵⁸⁶ Zachary K Rothschild and others, 'Displacing blame over the ingroup's harming of a disadvantaged group can fuel moral outrage at a third-party scapegoat' (2013) 49 *J Experimental Social Psychology* 898, 898

⁵⁸⁷ Taylor and Moghaddam (n 547) 78-80

⁵⁸⁸ *ibid* 79

⁵⁸⁹ Tajfel and others (n 569) 153

⁵⁹⁰ Bethany Lassetter and Rebecca Neel, 'Malleable liberals and fixed conservatives? Political orientation shapes perceived ability to change' (2019) 82 *J Experimental Social Psychology* 141

⁵⁹¹ Discussed above at text accompanying (n 554) to (n 556); see also T. F. Pettigrew, 'The ultimate attribution error: Extending Allport's cognitive analysis of prejudice' (1979) 5 *Personality and Social Psychology Bulletin* 461

⁵⁹² Taylor and Moghaddam (n 547) 168

torture as more morally justified when done by their own nation's security services than by the other nation's.⁵⁹³

Tajfel further describes social comparison in terms of an individual's personal comparison with his or her ingroup or outgroup.⁵⁹⁴ Many studies have shown how social comparison affects the self-concept, both in terms of personal identity or self-esteem as delineated by Taylor.⁵⁹⁵ Where those comparisons lead to positive or negative assessments of our own group, IET suggests we will experience group-specific emotional responses as a result.⁵⁹⁶ Costarelli and Colis reasoned that ambivalent assessments of the ingroup may produce different outcomes for people who identify with the group in a more evaluative or emotional sense compared to those who identify in a more descriptive or cognitive sense.⁵⁹⁷

Social comparison may also send us to the mental 'toolbox' to do some self-esteem maintenance when our self-concept is threatened. Crawford found that participants tended to mentally distance themselves from a competitor who out-performed them when task performance was relevant to the participants' sense of 'individual' self, emphasizing the other's non-shared identity, even though the competitor exhibited both ingroup and outgroup characteristics. But when the task was not self-relevant, "participants were much more likely to focus on the shared identity as a way to increase perceived similarity between the self and the other as a method of maintaining positive self-evaluation through reflection" as a fellow group member.⁵⁹⁸ Other research suggests that in addition to the self-enhancing effects observed in personal evaluations relative to outgroup members, we may tend to be more self-critical when comparing ourselves to ingroup members.⁵⁹⁹ At the same time, external explanations for group failure may tend to protect the self while maintaining group cohesion.⁶⁰⁰ Integrating a more psychodynamic approach, Cichocka has distinguished two kinds of in-group positivity: a secure type deriving from personal fulfilment as a group

⁵⁹³ Mark Tarrant and others, 'Social identity and perceptions of torture: It's moral when we do it' (2012) 48 *J Experimental Social Psychology* 513

⁵⁹⁴ Tajfel (n 570) 11

⁵⁹⁵ Discussed above at text accompanying (n 578) to (n 582)

⁵⁹⁶ Discussed above at text accompanying (n 583) to (n 586)

⁵⁹⁷ Sandro Costarelli and Elena Colis, 'In-group Attachment and Glorification, Perceptions of Cognition-based ambivalence as Contributing to the Group, and Positive Affect' (2016) 24 *Current Research in Social Psychology* 31, 32; see also Sandro Costarelli, 'The Pros and Cons of Ingroup Ambivalence: The Moderating Roles of Attitudinal Basis and Individual Differences in Ingroup Attachment and Glorification' (2015) 23 *Current Research in Social Psychology* 26

⁵⁹⁸ Matthew T Crawford, 'The renegotiation of social identities in response to a threat to self-evaluation maintenance' (2007) 43 *J Experimental Social Psychology* 39, 45-46

⁵⁹⁹ Michael R Ransom, Chris Kast, and Robert K Shelly, 'Self-Enhancement, Self-Protection and Ingroup Bias' (2015) 23 *Current Research in Social Psychology* 56, 62

⁶⁰⁰ Taylor and Moghaddam (n 547) 168-69

member and a defensive type deriving from unsatisfied individual needs—the latter corresponding with “collective narcissism” as a compensatory tool.⁶⁰¹

In addition to the drive for a positive group/self-evaluation as described in the preceding examples, Tajfel also points to psychological distinctiveness as a motivating factor that arises from social identity.⁶⁰² Psychological distinctiveness refers to our drive to find a ‘niche’ identity that shows our group in a “distinct and positive light.”⁶⁰³ Football clubs all basically do the same thing (kick a ball up and down the pitch), but distinguish each other through team colours, a regional home base, specific chants or music, etc., that make each stand out from the other. One way social identification may drive behaviour is by reducing our subjective uncertainty about who we are.⁶⁰⁴ The more uncertain we are about our own distinctiveness, the more we may rely on identification with our social group to fill in the gaps in our self-concept.⁶⁰⁵

In the context of political engagement, ideological uncertainty (in other words, not identifying strongly with a political ingroup) may actually undermine our sense of political efficacy and interest in politics, thereby motivating individuals to disengage and withdraw from participating in electoral politics rather than engaging more deeply.⁶⁰⁶ Positive aspects of our ingroup, however, such as the group’s perceived ability to deal effectively with threats, can drive ingroup identification.⁶⁰⁷ Even though Sim and others tend to favour a more stable, individuated structuring of the self, their research found that when ingroups with strong relevance to a participant’s self-concept were made salient, participants increasingly endorsed ingroup traits as self-descriptive (complementing rather than displacing the “core” self).⁶⁰⁸ Thinking about our social groups can even influence our eating behaviour⁶⁰⁹ and

⁶⁰¹ Aleksandra Cichocka, ‘Understanding defensive and secure in-group positivity: The role of collective narcissism’ (2016) 27(1) *European Review of Social Psychology* 283, 287, 305-06; Irem Eker, Aleksandra Cichocka, and Chris G Sibley, ‘Investigating motivations underlying collective narcissism and in-group identification’ (2023) 26(3) *Group Processes & Intergroup Relations* 669; see also Taylor and Moghaddam (n 547) 107-108 (the collective narcissism dynamic appears to comport with theories of intergroup relations that stress distinctions in comparative processes between advantaged and disadvantaged groups like equity theory—in that sense as a means for disadvantaged group members to restore ‘psychological equity’ by aggrandizing the group and in contrast to leaving the group to seek satisfaction elsewhere)

⁶⁰² Tajfel (n 570) 24-27

⁶⁰³ Taylor and Moghaddam (n 547) 80-81

⁶⁰⁴ Michael A Hogg, ‘Subjective Uncertainty Reduction through Self-categorization: A Motivational Theory of Social Identity Processes’ in Wolfgang Stroebe and Miles Hewstone (eds), 11 *European Review of Social Psychology* (Wiley 2000) 223

⁶⁰⁵ Fiona Grant and Michael A Hogg, ‘Self-uncertainty, social identity prominence and group identification’ (2012) 38 *J Experimental Social Psychology* 538

⁶⁰⁶ Joseph A Vitriol and others, ‘Ideological uncertainty and investment of the self in politics’ (2019) 82 *J Experimental Social Psychology* 85

⁶⁰⁷ Stefano Pagliaro and others, ‘Group based resiliency: Contrasting the negative effects of threat to the ingroup’ (2013) 21 *Current Research in Social Psychology* 35

⁶⁰⁸ Jessica J Sim and others, ‘How social identity shapes the working self-concept’ (2014) 55 *J Experimental Social Psychology* 271, 276; see also R C Morris, ‘Identity Salience and Identity Importance in Identity

actually affects how tasty we find particular foods⁶¹⁰ (*e.g.*, Americans and the English both like beans, both like toast, but only the English would put beans *on* toast and think, “yummy breakfast”). Each of these cases ultimately emphasizes the importance of group identity in assembling a picture of ourselves that incorporates group values and norms as our own.

When we highly identify with our ingroup, presence of an undesirable ingroup member or “black sheep” who deviates from our picture of the group can promote more positive evaluations of our stereotype of the group and motivate more negative evaluations of the black sheep.⁶¹¹ However, Van Swol found that the presence of an “extreme” member of the group (who essentially over-represents the views of the group) may not significantly contribute to the polarization of the group.⁶¹² Van Swol’s finding runs counter to the expectation that “based on social comparison and persuasive arguments, groups containing an extreme member should polarize more,”⁶¹³ and may reflect subsequent findings that suggest the ways in which groups form can influence the group’s tolerance of deviant opinion expression. Jans and others found that inductively formed groups, established by idiosyncratic contributions (like the small, ad hoc discussion groups in Van Swol’s experiment), were more tolerant of deviant opinion than groups where social identity is deduced from ingroup similarities.⁶¹⁴ The porous nature of online groups on social media (excepting private chats or messages) means that any and all of these tendencies may play out quite frequently, demanding careful scrutiny of how online groups assemble and self-regulate.

Alternative accounts look to demonstrate links between other “psychological needs, motives, and properties” and the common adoption of particular political ideologies by characteristically similar individuals.⁶¹⁵ Jost and others state that the ‘motivated social-cognitive’ perspective assumes that individuals adopt ideologies “in part because they satisfy some psychological needs.”⁶¹⁶ Such needs range from personal desires for self-enhancement

Theory’ (2013) 21 *Current Research in Social Psychology* 23 (stressing the distinction between identity salience and psychological centrality, or importance to self-concept)

⁶⁰⁹ Jinyu Liu, Jason Michael Thomas, and Suzanne Higgs, ‘The relationship between social identity, descriptive social norms and eating intentions and behaviors’ (2019) 82 *J Experimental Social Psychology* 217

⁶¹⁰ Leor M Hackel and others, ‘From groups to grits: Social identity shapes evaluations of food pleasantness’ (2018) 74 *J Experimental Social Psychology* 270

⁶¹¹ Paul Hutchison and others, ‘Getting rid of the bad ones: The relationship between group identification, deviant derogation, and identity maintenance’ (2008) 44 *J Experimental Social Psychology* 874, 880

⁶¹² Lyn M Van Swol, ‘Extreme members and group polarization’ (2009) 4 (3) *Social Influence* 185, 194

⁶¹³ *ibid* 189

⁶¹⁴ Lise Jans and others, ‘Dynamic Reactions to opinion deviance: The role of social identity formation’ (2019) 84 *J Experimental Social Psychology* 103803

⁶¹⁵ See, *e.g.*, John T Jost and others, ‘Political Conservatism as Motivated Social Cognition’ (2003) 129 (3) *Psychological Bulletin* 339, 339

⁶¹⁶ *ibid* 340

or cognitive ‘closure’ and uncertainty avoidance to threat management and preferences about social change or adherence to social norms.⁶¹⁷ Many studies of motivated social cognition accordingly seek to quantify the association between specific psychological needs and political ideologies. Strupp-Levitsky and others, for example, posit the association they found between ‘epistemic’ motives like need for cognitive closure, ‘binding’ moral foundations like deference to authority, and participants’ self-identified conservatism suggest conservatism (as an ideology) is associated with qualitatively different psychological concerns from liberalism—finding distinct associations between empathetic motivation, ‘individualizing’ moral foundations like harm avoidance and participants’ self-identified liberalism.⁶¹⁸ Jost holds that the disparate associations shown in such motivated social cognition studies shed essential psychological light on ideological asymmetries: political differences related to particular affinities between discrete cognitive-motivational processes and the content of specific belief systems.⁶¹⁹

Jost’s framework specifically hypothesizes a link between “specific social, cognitive, and motivational needs” and the “ideological core of political conservatism”—building on preceding sociological accounts of conservative ideology.⁶²⁰ Jost and others incorporate various measures of selected “core aspects” of conservative ideology, including Altemeyer’s Right-Wing Authoritarianism scale (“RWA”) (as a representation of resistance to socio-cultural change) and Pratto and others’ Social Dominance Orientation scale (“SDO”) (as a representation of acceptance of inequality),⁶²¹ RWA theory proposes authoritarianism as a set of fearful and aggressive responses to situational threats (social, economic, and/or political), which in turn have been correlated to a broad range of attitudes and behaviours.⁶²² SDO theory focuses on biological and social factors rather than affective differences, proposing that universal drives to minimize group conflict promote the development of justificatory belief systems, including those that preserve existing hierarchies.⁶²³ Jost and

⁶¹⁷ John T Jost, Brian A Nosek, and Samuel D Gosling, ‘Ideology: Its Resurgence in Social, Personality, and Political Psychology’ (2008) 3 (2) *Perspectives on Psychological Science* 126, 128-133

⁶¹⁸ Michael Strupp-Levitsky and others, ‘Moral “foundations” as the product of motivated social cognition: Empathy and other psychological underpinnings of ideological divergence in “individualizing” and “binding” concerns’ (2020) 15 (11) *PLoS ONE*: e0241144 <<https://doi.org/10.1371/journal.pone.0241144>> accessed 15 October 2023

⁶¹⁹ John T Jost, ‘Ideological Asymmetries and the Essence of Political Psychology’ (2017) 38 (2) 167, 168-71; *cf.* Timothy P Collins, Jarret T Crawford, and Mark J Brandt, ‘No evidence for ideological asymmetry in dissonance avoidance: Unsuccessful close and conceptual replications of Nam, Jost, and Van Bavel’ (2013) 48 (3) *Social Psychology* 123 (finding that liberals and conservatives were equally likely to avoid writing counter-attitudinal essays—a test aimed at arousing cognitive dissonance)

⁶²⁰ Jost and others (n 615) 342

⁶²¹ *ibid*

⁶²² *ibid* 345-46

⁶²³ *ibid* 349-50

others observe that SDO scores nevertheless correlate with particular political attitudes, though noting that RWA and SDO may be motivated by “somewhat different concerns.”⁶²⁴

Other studies have found that personality or trait differences do not always translate into asymmetries in political behaviour.⁶²⁵ Casey, Vanman, and Barlow investigated partisan differences in empathy in an intergroup context and found that both liberal and conservative partisans had less empathy toward their political outgroup than ingroup or a ‘neutral’ group,⁶²⁶ despite previous findings that conservatives were generally less empathetic than liberals.⁶²⁷ The importance of intergroup context may also play a role in reconciling contradictory findings⁶²⁸ regarding asymmetries in information processing: Kahan’s study of decision making about political group-relevant empirical issues (such as climate change) found that “when assessing evidence of the other group’s propensity to consider evidence in an open-minded and reflective way, liberals and conservatives were *uniformly* prone to ideologically motivated reasoning.”⁶²⁹ Zmigrod has proposed a structural framework for studying ideologies (regardless of their content or aim)⁶³⁰ by focusing on descriptive and prescriptive ‘doctrinal’ components⁶³¹ alongside ingroup and outgroup ‘relational’ components.⁶³² Zmigrod contends this integrated approach could effectively capture both “the nuanced processes that confer initial susceptibility to internalizing ideological doctrines, as well as the mechanisms that reinforce or dampen these effects”⁶³³—in other words, getting a clearer picture of the “where, when, and why” of ideological thinking⁶³⁴ by

⁶²⁴ Jost and others (n 615) 350

⁶²⁵ See, e.g., James P Casey, Eric J Vanman, and Fiona Kate Barlow, ‘Empathetic Conservatives and Moralizing Liberals: Political Intergroup Empathy Varies by Political Ideology and is Explained by Moral Judgment’ (2023) 0 (0) *Personality and Social Psychology Bulletin* 1 <<https://doi.org/10.1177/01461672231198001>> accessed 15 October 2023; see also Leor Zmigrod and Amit Goldenberg, ‘Cognition and Emotion in Extreme Political Action: Individual Differences and Dynamic Interactions’ (2021) 30 (3) *Current Directions in Psychological Science* 218 (examining the interaction between cognition and emotion relative to the likelihood of engaging in extreme ideological action)

⁶²⁶ Casey, Vanman, and Barlow (n 625) 8 (measuring empathetic concern, perspective-taking, and empathetic intentions)

⁶²⁷ *ibid* (finding liberal participants showed an empathy-mediating effect through outgroup moral judgments more strongly than conservative participants); *ibid* 19–21 (finding further that higher perceptions of outgroup harmfulness among liberal participants statistically explained the asymmetry in empathy)

⁶²⁸ Relative to those along the lines of Jost (n 619) 174–75 reporting an ideological difference in liberals’ higher self-assessments of need for cognitive deliberation

⁶²⁹ Dan M Kahan, ‘Ideology, motivated reasoning, and cognitive reflection’ (2013) 8 (4) *Judgment and Decision Making* 407, 417 (finding no correlation between participants’ self-identification with the Republican Party and scores on the Cognitive Reflection Test, a performance-based measure of the disposition to engage in conscious and effortful information processing)

⁶³⁰ Leor Zmigrod, ‘A Psychology of Ideology: Unpacking the Psychological Structure of Ideological Thinking’ (2022) 17 (4) *Perspectives on Psychological Science* 1072, 1075–76

⁶³¹ *ibid* 1075 (beliefs and ‘dogmas’ that explain the world around the adherent and provide moral or logical foundations for future action)

⁶³² *ibid* 1075–76 (demarcations of ingroup and outgroup identity, tendencies to favoritism or hostility)

⁶³³ *ibid* 1081

⁶³⁴ *ibid* 1082

recognizing social psychological phenomena that may mediate the responses and actions predicted by dispositional theories of political psychology.

The key social psychological concepts presented in this section included processes of social identification and several accounts of motivated decision making. Social categorization and comparison inform and shape the way we think about and believe in ourselves, others, and the world around us. Social group membership serves as an important cue for those processes. Social identification in turn influences behaviours that relate to group membership. Self-esteem, emotions, and other factors that motivate behaviour likewise have important 'group' components which reflect or interact with group membership. Additional concepts from political psychology may provide further insights into the nature of observed asymmetries between groups. In the next section, I discuss some of the studies concerning these concepts in specific media contexts.

The Social Psychology of Political News and Social Media

Social psychology concepts provide a useful lens for examining what happens when people encounter political news content on social media. Social identification processes play an important role when members of polarized groups take in a particular piece of information and helps explain why the same information may be assessed differently by those group members. Group membership in turn plays a role in motivating information seeking, sharing, discounting, and rejection behaviours. In addition, the medium itself may exacerbate tendencies in group identification due to the ease and ubiquity of enacting identity-related behaviours.

As discussed in Part One, the 'problem' of fake news encompasses a number of issues related to misinformation, abuse, public trust in the press, and participation in democracy.⁶³⁵ In the pre-Internet era, Lord, Ross, and Lepper questioned whether opposing factions in a political policy dispute (*e.g.*, capital punishment) would become more or less polarized when they took in mixed or inconclusive findings on the issue in dispute.⁶³⁶ Although the participants viewed the same dataset, the faction members came away with opposing impressions of the quality and persuasiveness of the data⁶³⁷ and reported greater attitude polarization than before encountering the information.⁶³⁸ The ease of finding and sharing stories on social media for or against any particular issue probably cannot be overstated, and

⁶³⁵ Part I, 'The Fake News Problem'

⁶³⁶ Charles G Lord, Lee Ross, and Mark R Lepper, 'Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence' (1979) 37 (11) *J of Personality and Social Psychology* 2098, 2099

⁶³⁷ *ibid* 2101-02

⁶³⁸ *ibid* 2103-04

the biased assimilation phenomena observed by Lord, Ross, and Lepper are likely not unique to the media they studied. When we encounter political news online, we may be similarly likely to accept news that supports our own viewpoints while tending to scrutinize opposing viewpoints. Social media provides new and additional contexts, mediums, and means for interaction which intensify the “hostile media phenomenon” described by Vallone, Ross, and Lepper in 1985.⁶³⁹ They found that partisan viewers of U.S. network television coverage of the Beirut massacre each rated programs as biased against their particular political group, reporting opposing assessments of fairness and divergent recollections of the actual contents of the news programs.⁶⁴⁰

Social media makes the social context of encountering news omnipresent in the user experience—stories there are necessarily shared and understood by and among people in public social relation to one another, including political opposition or hostility. Guess, Nyhan, and Reifler tracked users’ web browsing to examine the role of Facebook in distributing “fake news” during the final weeks of 2016 U.S. Presidential election, finding evidence of Americans’ selective exposure to misinformation based on candidate preference, particularly among conservative partisans.⁶⁴¹ Assessing the relative importance of biased information processing and partisan ‘cheerleading’⁶⁴² as explanations for misinformation, Peterson and Iyengar probed how the availability of incentives for accuracy might change the typical news media diet of political partisans.⁶⁴³ They found no effect of the incentives on news outlet choice⁶⁴⁴ and argue their finding suggests “that the partisans who use them do not believe the information provided by party-congenial sources is any less accurate than information from nonpartisan or expert sources.”⁶⁴⁵ People also appear to be motivated to selectively seek confirmatory information after making predictions (of varying degrees of arbitrariness) in anticipation of the positive emotions associated with being correct.⁶⁴⁶ As discussed previously, emotion in social context can influence intergroup behaviour,⁶⁴⁷ and the

⁶³⁹ Robert P Vallone, Lee Ross, and Mark R Lepper, ‘The Hostile Media Phenomenon: Biased Perception and Perceptions of Media Bias in Coverage of the Beirut Massacre’ (1985) 49 (3) *J of Personality and Social Psychology* 577, 577-78

⁶⁴⁰ *ibid* 579-82

⁶⁴¹ Andrew Guess, Brendan Nyhan, and Jason Reifler, ‘Selective Exposure to Misinformation: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign’ (9 January 2018) 1, 19 <<https://www.dartmouth.edu/~nyhan/fake-news-2016.pdf>> accessed 02 June 2020

⁶⁴² Essentially, knowingly disingenuous responses to information which serve to manage or bolster impressions of the partisan’s ingroup

⁶⁴³ Erik Peterson and Shanto Iyengar, ‘Partisan Gaps in Political Information and Information-Seeking Behavior: Motivated Reasoning or Cheerleading?’ (2021) 65 (1) *Am J Pol Sci* 133

⁶⁴⁴ *ibid* 141-45 (further finding that participants behaved similarly in ‘real world’ web browsing, while also noting challenges in comparing study incentives to actual social pressures or benefits regarding accuracy)

⁶⁴⁵ *ibid* 142

⁶⁴⁶ Aaron M Scherer, Paul D Windschitl, and Andrew R Smith, ‘Hope to be right: Biased information seeking following arbitrary and informed predictions’ (2013) 49 *J Experimental Social Psychology* 106

⁶⁴⁷ *e.g.*, Mackie, Smith, and Ray (n 583)

architecture of social media—particularly recommendation algorithms—appears to be highly attuned to encouraging users’ emotional engagement with content on the platforms, as indicated by Zuckerberg’s comments on the way Facebook leverages users’ emotional reactions to News Feed items.⁶⁴⁸

The political utility of social media, both to the users and to the political class, has been touted since its early days,⁶⁴⁹ and accordingly, social media’s operationalization in politics sharpens contextual focus on political and ideological groups in particular. During the 2018 Irish abortion referendum, a study by Murphy and others found that participants exhibited an increased susceptibility to forming false memories for fake news that aligned with their political beliefs.⁶⁵⁰ Klemfuss and others looked at the relative amount of misinformation in a narrative and perceptions of news source credibility to see how they mediated memory accuracy.⁶⁵¹ They found that exposure to more misinformation hurt memory performance, generally speaking, but also found a “dampening effect” as study participants became skeptical of increasingly misinformation-heavy narratives.⁶⁵² However, where their participants thought the narrative was more credible, they tended to be less careful about scrutinizing information sources for accuracy.⁶⁵³ Because social media platforms are engineered around the idea of maximizing user engagement,⁶⁵⁴ conceptualizing the impact of these effects in the wild may require keeping in mind that the misinformation tap can be always on and wide open compared to the drip-drip-drip pace of traditional media distribution.

Anti-vaccination content on social media, for example, ranges from Facebook posts and stories to full-on multimedia and video documentary series, employing a range of persuasion techniques as well as parasocial interactions (the kinds of bonds formed by online liking, commenting, and subscribing) with anti-vax movement personalities like Andrew Wakefield.⁶⁵⁵ Social media platforms dramatically lower the cost and extend the reach of such content-rich mixed-media communications. Chaiken and Eagly’s study of the effects of communication modality on persuasion is thus even more important now than in

⁶⁴⁸ See Part II, ‘Platform Policies and Technical Capabilities’ – Facebook

⁶⁴⁹ See generally Part I, ‘Social Media News Use’

⁶⁵⁰ Gillian Murphy and others, ‘False Memories for Fake News During Ireland’s Abortion Referendum’ (2019) 30 (10) *Psychological Science* 1449

⁶⁵¹ J Zoe Klemfuss and others, ‘The Effects of Exposure to Differing Amounts of Misinformation and Source Credibility Perception on Source Monitoring and Memory Accuracy’ (2017) 4 (4) *Psychology of Consciousness: Theory, Research, and Practice* 337

⁶⁵² *ibid* 342-43

⁶⁵³ *ibid* 342-44 (looking at source credibility as a potential mediator between misinformation and memory accuracy)

⁶⁵⁴ Discussed in Part I, ‘Social Media News Use’ and Part II, ‘Platform Policies and Technical Capabilities’

⁶⁵⁵ Amanda S Bradshaw and others, ‘Propagandizing anti-vaccination: Analysis of *Vaccines Revealed* documentary series’ (2020) 38 *Vaccine* 2058, 2060-63

the days of video and cassette tapes when their investigation was originally undertaken: audio-visual media make information about the communicator more vivid (through factors such as attractiveness or shared social identity), while written messages may better highlight the message content itself.⁶⁵⁶ Bizer, Larsen, and Petty highlighted another way presentation can influence persuasiveness, describing how negatively framing an attitude (e.g., opposition toward one political candidate rather than support for another) can enhance attitude certainty⁶⁵⁷—which in turn can reinforce attitude-consistent behavioural intentions.⁶⁵⁸ By presenting any and every mode of media engagement imaginable, fake news campaigns over social media have the potential to be powerfully persuasive.

Hornsey and others put together a study of 2016 U.S. Presidential election voters to examine the association between support for Trump and anti-vaccination sentiment.⁶⁵⁹ One perhaps obvious explanation for anti-vaxxers supporting Trump is that the other choice in the election, Democratic candidate Hillary Clinton, was essentially no choice at all to voters who see themselves first and foremost as anti-vaccination, regardless of any other liberal or conservative leanings they might have.⁶⁶⁰ Yet Hornsey and others decided to test the extent to which a different (but overlapping) group, Trump voters, conformed to Trump's expression of anti-vaccination sentiment.⁶⁶¹ Unsurprisingly, they found a correlation between Trump support and vaccine concern, but also determined that the tendency was not statistically significant when political conservatism and conspiracist ideation among Trump voters were accounted for as explanations for vaccine concern,⁶⁶² although such factors could contribute to anti-vax sentiment.

Their finding of a modest interaction between Trump's anti-vax tweets and participant vaccine concern (but not Trump's golf tweets), was said to reflect a difference between Trump voters and non-Trump voters⁶⁶³—again, failing to realize they had conflated anti-vax Trump voters with Trump voters writ large (and pro-vax non-Trump voters with

⁶⁵⁶ Shelly Chaiken and Alice H Eagly, 'Communication Modality as a Determinant of Persuasion: The Role of Communicator Salience' (1983) 45 (2) *J of Personality and Social Psychology* 241, 253-54

⁶⁵⁷ George Y Bizer, Jeff T Larsen, and Richard E Petty, 'Exploring the Valence-Framing Effect: Negative Framing Enhances Attitude Strength' (2011) 32 (1) *Political Psychology* 59, 64-65

⁶⁵⁸ *ibid* 75-76

⁶⁵⁹ Matthew J Hornsey and others, 'Donald Trump and vaccination: The effect of political identity, conspiracist ideation and presidential tweets on vaccine hesitancy' (2020) 88 *J Experimental Social Psychology* 103974, 1

⁶⁶⁰ Hornsey and others excluded from analysis participants who recorded that they did not vote—i.e., those who may have been anti-vax and also anti-Trump, but who nevertheless could not be pro-Clinton, who is pro-vax, (n 659) 3; similarly, conservative partisans may simply have sought out more candidate information *necessarily*, as Trump was a relative political newcomer, while Clinton has been a public political figure for several decades: *cf.* Guess, Nyhan, and Reifler (n 641) 17-18

⁶⁶¹ Hornsey and others (n 659) 3

⁶⁶² *ibid*

⁶⁶³ *ibid* 4, 6-7 (again, not including non-voters, so either Clinton or third party)

non-Trump voters writ large). As a consequence, they miss the more parsimonious possible explanation that for anti-vaxxers who saw Trump's anti-vax tweets, their *anti-vax* social identity was relevant to their subsequent attitude expressions, not their Trump-supporter social identity—after all, they did not become decidedly more pro-golf after seeing Trump's golf tweets,⁶⁶⁴ though of course 'golfer' likely wasn't a relevant group identity in this case. A subsequent follow-up study could examine whether these voters' support for Trump may have dissipated after Trump's change in stance on vaccinations ahead of the 2020 election. Due to the frequently political character of fake news, researchers must be exceedingly careful to check for the influence of their own categorization and construal of the social situations they create or encounter, even at the abstracted level of experiment and survey data.

Social psychology also offers some specific insights into the decision making processes that play a role in fake news sharing behaviours. Dhir and others found positive associations between sharing fake news and ostensibly positive aspects of social media use, such as self-disclosure and online trust which are intrinsic to many of the platforms.⁶⁶⁵ How people draw different moral judgments regarding "loyal lies" and "disloyal honesty" may also influence a person's determination of whether sharing a fake story would be unethical—a judgment that tends to be different for oneself than for others.⁶⁶⁶ Simply encountering misinformation repeatedly has also been found to make spreading fake stories seem less unethical, regardless of whether the misinformation is believed.⁶⁶⁷ This is another aspect where the volume of content and ease of sharing it over social media likely exacerbates such tendencies.

This may play a role in the phenomenon media studies have described as the supposed 'filter bubble' occurring on social media platforms like Twitter.⁶⁶⁸ Consider, for instance, the #tcot "top conservatives on Twitter" group: evidence suggests a tweeter becomes a #tcot less through interaction or debate with ideologically opposed liberals than

⁶⁶⁴ Hornsey and others (n 659) 6-7

⁶⁶⁵ Amandeep Dhir and others, 'Why do people share fake news? Associations between the dark side of social media use and fake news sharing behavior' (2019) 51 J Retailing and Consumer Services 72

⁶⁶⁶ John A D Hildreth and Cameron Anderson, 'Does loyalty trump honesty? Moral judgements of loyalty-driven deceit' (2018) 79 J Experimental Social Psychology 87; cf. Tessa L Dover and others, 'Does unfairness feel different if it can be linked to group membership? Cognitive, affective, behavioral and physiological implications of discrimination and unfairness' (2014) 56 J Experimental Social Psychology 96 (unfair treatment elicits different responses depending on group context, i.e., within groups versus between groups)

⁶⁶⁷ Daniel A Effron and Medha Raj, 'Misinformation and Morality: Encountering Fake-News Headlines Makes Them Seem Less Unethical to Publish and Share' (2020) 31 (1) Psychological Science 75

⁶⁶⁸ Part I, 'Fake News Studies'

by tweeting and re-tweeting to the choir of fellow conservatives.⁶⁶⁹ A study by Travaglino and others suggested that individuals can simultaneously engage with online activity in problematic ways while at the same time deriving a social identity (in this case, as a 'gamer') that provides emotional (and even material) support for its members in ways that their offline identities (like 'student') may not.⁶⁷⁰ To the extent that sharing and circulating informational content online can build communities of likeminded individuals, some of the self-categorization and self-esteem maintenance phenomena discussed earlier may play a part in motivating specific fake news interactions.

Each of the studies presented in this section highlight the role social identification plays in our encounters with media. In particular, political group identification affects our interactions with media content concerning our own political groups and of our civic rivals. As we seek out or reject information that touches on our group identity, the psychological processes at play may lead us to be selective or even less scrupulous about whether the information is genuine or fake. When we engage in these behaviours on social media, the architecture of the platforms in turn may push us further into biased or polarized ways of relating to one another each time we decide to 'like' or share a group-relevant story. While our usual mode of satisfying the drive for a positive, distinct group identity as discussed above does not necessarily encourage us to be particularly discerning between genuine and fake news accounts, social media companies are increasingly asking users to do just that through content flagging initiatives. Twitter has recently added functionality for users in Australia, Brazil, Philippines, South Korea, Spain, and the U.S. to report tweets for misleading information, for example.⁶⁷¹ In the next section, I discuss implications of social psychology research for social media legal policy as well as the formulation of my research question.

Constraints on Content Moderation Policy Effectiveness

The preceding discussions brook a number of questions for strategically addressing the problem of fake news on social media platforms. First, we can ask how the status quo reflects the operation of social psychological principles. This may allow for a more complete diagnosis of the fake news problem by accurately describing what is going on with users when they encounter political news content on social media. Second, we can ask how certain interventions operate with regard to group identification. This could help explain why efforts like media literacy campaigns may tend to produce mixed results. Third, we can ask

⁶⁶⁹ M D Conover and others, 'Political Polarization on Twitter' (2011) Proceedings of the Fifth International Association for the Advancement of Artificial Intelligence Conference on Weblogs and Social Media 89

⁶⁷⁰ Giovanni A Travaglino and others, 'We are all in this together: The role of individuals' social identities in problematic engagement with video games and the internet' (2020) 59 (2) British J Social Psychology 522

⁶⁷¹ Twitter Help Center, 'Report violations' (2023) <<https://help.twitter.com/en/rules-and-policies/twitter-report-violation>> last accessed 11 March 2023

how policy proposals likely will perform in light of the constraints imposed by the psychological processes at play. Answering this question serves as the basis for my empirical research discussed in the next Part.

The Status Quo

Fake news has been described as a matter of international concern, interfering with “the public’s right to know and the right of individuals to seek and receive, as well as to impart, information of all kinds.”⁶⁷² But as discussed in Part One, members of the global public have increasingly come to rely upon or even prefer social media platforms as their source of news information.⁶⁷³ Accordingly, social media platforms like Twitter, Facebook, and YouTube promote or feature news content in order to attract and retain users, whom the platforms package and offer to paying advertisers. Likewise, news producers that rely on advertising revenue create attention-grabbing, click-inducing content to attract views. The alignment of the platforms’ and content providers’ incentive structures to bring in maximal audiences thus points towards a more permissive media environment in terms of both subject matter and subjectivity. YouTube CEO Susan Wojcicki has expressed her commitment to keeping YouTube an open platform for user-generated content while “working really hard to figure out what’s the right way to balance responsibility with freedom of speech.”⁶⁷⁴ Platforms and governments will doubtlessly continue working hard to strike a principled balance between profitable and acceptable content.

In terms of the current law and regulatory regimes, governments mostly have left decisions about content to the social media platforms themselves, while at the same time providing the platforms with protection from liabilities arising due to content generated by users over their services, as discussed in Part Two.⁶⁷⁵ The social media platforms and news producers could conceivably remain agnostic as to the kind or quality of news content that users encounter, so long as their advertisers would pay for the audiences who engaged with that content. However, governments, along with other institutions and the users themselves, have a decided interest in what the news says. Indeed, governments might even take notice of what is promulgated over social media out of existential necessity: numerous examples in

⁶⁷² UN Special Rapporteur on Freedom of Opinion and Expression, OSCE Representative on Freedom of the Media, OAS Special Rapporteur on Freedom of Expression, and ACHPR Special Rapporteur on Freedom of Expression and Access to Information, ‘Joint Declaration on Freedom of Expression and “Fake News”, Disinformation and Propaganda’ FOM.GAL/3/17 (Vienna, 3 March 2017) <<https://www.ohchr.org/Documents/Issues/Expression/JointDeclaration3March2017.doc>>

⁶⁷³ Part I, ‘Social Media News Use’

⁶⁷⁴ Lesley Stahl, ‘How does YouTube handle the site’s misinformation, conspiracy theories, and hate?’ (60 Minutes, 01 December 2019) <<https://www.cbsnews.com/news/is-youtube-doing-enough-to-fight-hate-speech-and-conspiracy-theories-60-minutes-2019-12-01/>> accessed 05 June 2020

⁶⁷⁵ Part II, ‘Social Media and Media Systems’

the past twenty years, particularly amidst popular uprisings under authoritarian regimes, highlight the role of social media in political movements and the organization (and re-organization) of civil society.⁶⁷⁶

This suggests that the tendencies toward biased or selective information seeking and sharing over social media discussed in this Part will continue. The vast availability of political news content online means that users may encounter a near-limitless amount of information that could bear positively or negatively on the political groups with which they identify. Even where social media users interact behind screen-names and avatars, we still tend to ‘recognize’ others who share our ethnic and social groups.⁶⁷⁷ Furthermore, emotional engagement with online content is an important part of how social media platforms build and retain audiences, which likewise implicates intergroup relations in user interactions with content. Social media as a medium can engage with users through audio, video, and text—providing a rich multi-media experience but also a high potential level of persuasiveness. The problem of ‘fake news’ on social media thus involves more than just news content which isn’t genuine, but an array of connected psychological processes that may be inseparable from the way social media operates generally.

This helps explain why developing effective content moderation has proven so challenging for large social media platforms. Media researchers have argued that “[t]he words we choose to describe media manipulation can lead to assumptions about how information spreads, who spreads it, and who receives it . . . [shaping] what kinds of interventions or solutions seem desirable, appropriate, or even possible;”⁶⁷⁸ while econometricians have developed definitions of fake news that can be swallowed whole by arbitrary exceptions like “false statements by politicians” or misleading reports that are not “outright” false.⁶⁷⁹ But for users, calling certain content ‘fake news’ is to an extent an expression of anxiety and discomfort about what that content says about the social groups with which the users identify. This is a characteristic that cannot be defined with reference to the content alone: it arises only in the interaction between the information and the user. Despite the vast amounts of data that the social media platforms can generate about the lexical features of news content or the relationships between users,⁶⁸⁰ none of the moderation techniques described in Part Two attempt to implement an algorithm that could

⁶⁷⁶ Discussed in Part I, ‘Social Media News Use’ and Part I, ‘The Fake News Problem’

⁶⁷⁷ Devin G Ray and Christina Matschke, ‘Cross-group recognition bias generalizes to diverse non-face representations of digital identity’ (2012) 48 J Experimental Social Psychology 387

⁶⁷⁸ Caroline Jack, ‘Lexicon of Lies: Terms for Problematic Information’ (*Data & Society Research Institute*, August 2017) 1 <<https://datasociety.net/library/lexicon-of-lies/>> accessed 05 June 2020

⁶⁷⁹ Hunt Allcott and Matthew Gentzkow, ‘Social Media and Fake News in the 2016 Election’ (2017) 31(2) J Econ Perspectives 211, 213-14

⁶⁸⁰ Part II, ‘Platform Policies and Technical Capabilities’ – Evaluating Content Moderation at Scale

accurately anticipate the extent to which these psychological processes operate in a particular user with regard to a specific story.

Interventions

On a practical level, platforms approach the problem of fake news through their terms-of-service policies and technical solutions regarding content such as automating content moderation through artificial intelligence.⁶⁸¹ These tools are supplemented with data provided by users and expert sources like fact-checkers.⁶⁸² Perversely, some of the user-centred strategies for policing content, such as the tools provided by social media platforms to flag or report objectionable material, can simultaneously result in user empowerment with regard to content *and* in negative user outcomes like harassment⁶⁸³ and radicalization.⁶⁸⁴ Observations like these likely reflect some of the attitude polarization processes identified by Lord, Ross, and Lepper as discussed above.⁶⁸⁵ The findings of subsequent studies likewise suggest that media literacy campaigns and education programs to combat fake news likely may not succeed. Increased awareness and emotional involvement with a divisive political issue may simply result in increased bias.⁶⁸⁶

Biased assimilation similarly portends the fruitlessness of fact-checking schemes: “Individuals will dismiss and discount empirical evidence that contradicts their initial views but will derive support from evidence, of no greater probativeness, that seems consistent with their views.”⁶⁸⁷ Warnings about possible misinformation similarly may be of limited effectiveness.⁶⁸⁸ Furthermore, beliefs can survive the total discrediting of their initial evidential base, not due to any dysfunction or irrationality on the part of the believer, but simply due to the importance of context in belief formation.⁶⁸⁹ One of the early investigations of resisting persuasion by McGuire attempted “inoculation” against future persuasion by pre-exposing participants to weak counter-arguments against prevailing

⁶⁸¹ See, e.g., United Nations General Assembly ‘Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression’ (6 April 2018) UN Doc A/HRC/38/35

⁶⁸² Discussed in Part II, ‘Platform Policies and Technical Capabilities’

⁶⁸³ Kate Crawford and Tarleton Gillespie, ‘What is a flag for? Social media reporting tools and the vocabulary of complaint’ (2016) 18 *new media & society* 410

⁶⁸⁴ danah boyd, ‘You Think You Want Media Literacy... Do You?’ (*Data & Society: Points*, 9 March 2018) <<https://points.datasociety.net/you-think-you-want-media-literacy-do-you-7cad6af18ec2>> accessed 05 June 2020

⁶⁸⁵ Discussed *supra* (n 636) to (n 638)

⁶⁸⁶ Vallone, Ross, and Lepper (n 639) 577-78

⁶⁸⁷ Lord, Ross, and Lepper (n 636) 2099

⁶⁸⁸ Murphy and others (n 650) 1455-56; see also Peterson and Iyengar (n 643)

⁶⁸⁹ Craig A Anderson, Mark R Lepper, and Lee Ross, ‘Perseverance of Social Theories: The Role of Explanation in the Persistence of Discredited Information’ (1980) 39(6) *J Personality and Social Psychology* 1037

cultural truisms—mimicking the vaccination process.⁶⁹⁰ Cameron, Jacks, and O'Brien updated McGuire's inoculation procedure⁶⁹¹ with a battery of different strategies for resisting persuasion.⁶⁹² Their experiment tested direct counter-arguing; bolstering one's own position, and attacking the source of the persuasive message, among other techniques.⁶⁹³ Unlike McGuire,⁶⁹⁴ Cameron, Jacks, and O'Brien did not initially find that counter-arguing conferred effective resistance to persuasion, possibly because the participants were provided with prefabricated counterarguments rather than encouraged to develop their own.⁶⁹⁵ Their later testing involving idiosyncratic counter-arguments and attitude bolstering did find improved resistance to attitude change.⁶⁹⁶ Source derogation (a tactic routinely employed by today's online fact-checkers⁶⁹⁷) was not found to be immediately effective, and moreover was disfavoured by participants as a potential strategy due to its social undesirability.⁶⁹⁸

Even successfully resisting persuasion may subtly reduce the strength of the resister's targeted attitude and increase subsequent vulnerability.⁶⁹⁹ Tormala, Clarkson, and Petty postulated an interaction between resistance to persuasion and feedback on their own "resistance performance."⁷⁰⁰ However, an unfavourable assessment of resistance—even based on resisting "the same message in the same way and with the same objective success"—may leave people with less certainty about their own position.⁷⁰¹ Findings like these point in a different direction than the one urged by media literacy campaigners and fact-checkers. Users might stand a better chance of rationally assessing the news if they were more conscious of their tendencies to interpret information as social beings, but as discussed in Part Two, social media companies more typically engage with the pre-packaged solutions offered by media experts.⁷⁰²

⁶⁹⁰ William J McGuire, 'Resistance to Persuasion Conferred by Active and Passive Prior Refutation of the Same and Alternative Counterarguments' (1961) 63 (2) *J of Abnormal and Social Psychology* 326

⁶⁹¹ *ibid* 326

⁶⁹² Kimberly A Cameron, Julia Zuwerink Jacks, and Maureen E O'Brien, 'An Experimental Examination of Strategies for Resisting Persuasion' (2002) 7 *Current Research in Social Psychology* 205

⁶⁹³ *ibid* 206, 209-10

⁶⁹⁴ McGuire (n 690) 329-30

⁶⁹⁵ Cameron, Jacks, and O'Brien (n 692) 212

⁶⁹⁶ *ibid* 214

⁶⁹⁷ Discussed in Part II, 'Platform Policies and Technical Capabilities'

⁶⁹⁸ Cameron, Jacks, and O'Brien (n 692) 213, 214

⁶⁹⁹ Zakary L Tormala, Joshua J Clarkson, and Richard E Petty, 'Resisting Persuasion by the Skin of One's Teeth: The Hidden Success of Resisted Persuasive Messages' (2006) 91 *J Personality and Social Psychology* 423

⁷⁰⁰ *ibid* 423-24

⁷⁰¹ *ibid* 424, 432-34

⁷⁰² Part II, 'Platform Policies and Technical Capabilities'

Understanding Constraints

Just as we consider psychology in the assessment of legal evidence quality, we should consider social psychology in the assessment of social media ‘evidence’ concerning fake news. I contend that the key to understanding the difficulty of regulating fake news on social media is its acutely *social* experience. The users, the content moderators and algorithm programmers, journalists and members of media watchdog organizations, CEOs of the platforms, and lawmakers all encounter the news content on the platforms as members of social groups and in relation to other social groups. The ultimate value of a social psychological perspective on the fake news problem lies in its revelation of pitfalls and traps awaiting those who approach the issue with a purely technical, material approach to addressing social media’s role in the matter. Taylor and Moghaddam summarize social identity theory as follows:

The important motivation for group categorization is social identity. It is postulated that the need for social identity leads to a striving for group distinctiveness along positively valued dimensions. Group distinctiveness refers to those characteristics of a group that make it different from other groups—in other words, a stereotype. According to social identity theory, then, groups strive for a distinctive own-group stereotype that is valued positively by other groups. Thus, the creation and maintenance of a positively valued stereotype are postulated to be a fundamental human need.⁷⁰³

The insights on how these processes shape and drive intergroup behaviour⁷⁰⁴ suggest that unwary regulatory proposals may leave unaddressed many of the deepest challenges posed by fake news. Social psychological research has indicated the value of a number of research methodologies to usefully evaluate the pervasive influence of social context. The way forward for lawyers, academics, and lawmakers requires close attention to the processes of social identification and their connection to decision-making behaviour as well as other influences on intergroup relations that reinforce fake news and frustrate efforts to address it.

Past advances in communications technology demanded their own advances in legal ‘technology’ to meet new challenges arising from lower distribution costs, new media forms, and increased access to information. These advancements in the law have traditionally been tempered by fundamental values of free expression: The law has long been able to handle cases of fraud and defamation while safeguarding free speech and freedom of the press

⁷⁰³ Taylor and Moghaddam (n 547) 166

⁷⁰⁴ Discussed *supra*, ‘Principles of Social Psychology’ – Social Identity Theory

through due process. Where public policy obligates the social media companies to gather, assess, and act on information about user-generated content—essentially substituting regulatory and technical processes for the judicial process, the quality of data available to the companies should be an area of key concern, especially when legal consequences are attached to the companies' determinations. If policymakers' picture of the major platforms' techniques is in this sense too rosy, then the potential arises for the enshrinement of processes for judging speech that fundamentally rely on obscure or inapt criteria (relative to how we have evaluated speech traditionally). In order to gauge whether these new processes will have the quality necessary to reach just outcomes concerning news content on social media, I designed and conducted the experiment detailed in Part Four of this Thesis. In that next Part, I describe how I evaluated associations with social identity in the kinds of data social media companies may gather and rely upon, the methodological and ethical considerations I followed, and the results of my empirical study.

Key Takeaways

- By investigating the mental processes involved in intergroup relations, psychology helps explain decision making in social contexts like social media news use
- News about political groups can engage social media users through their social identity—that part of their sense of self tied to belonging to a political group
- Algorithmic moderation produces outcomes for speech based on various inputs, but those inputs may reflect motivations (like social identity) that bear on far more than the authenticity or accuracy of the content

Part Four: Experiment Design and Results

The social psychology research reviewed in Part Three suggested to me that social media users respond to news content not merely on its face but in relation to what the news implies about the social groups with which the users identify. In this fourth Part, I discuss the design and results of the experiment I conducted to investigate how political group identification relates to social media users' reactions to political news. In the first section, I discuss the methodological and ethical considerations I undertook in designing my experiment. My design aimed to measure participants' level of identification with two American political groups, the Republicans and Democrats, and then to capture their reactions to genuine and fake news content they might see on social media. In the second section, I report my findings. The analysis of my findings showed that in both political groups I studied, highly identified participants were more willing to flag genuine stories that denigrated their own group as fakes.⁷⁰⁵ The final section of this Part discusses the implications of those findings for efforts to address fake news on social media through content moderation techniques that rely on user data input, as discussed in Part Two.

Methodology

As discussed in Part Three, group membership plays an important role in our sense of identity. It helps to tell us what we are like and to distinguish ourselves from others based on our social groups, and this social identification process results in thinking and feeling distinctly as a group member. We may act to protect our group's reputation, in turn protecting the sense of belonging and esteem we get from being a group member. We likewise act and interact in motivated ways when we communicate over social media. A retweeted story or a News Feed post can, in a brief span, say a lot about one group or another, and we may react in particular ways if it says something hurtful about our own group. In order to answer whether political group identification relates to the social media user decisions that drive content moderation processes, I developed the following hypotheses.

⁷⁰⁵ *Republicans*, Kendall Rank Correlation (tau-beta = 0.151, p = 0.016); *Democrats*, Kendall Rank Correlation (tau-beta = 0.124, p = 0.014); see **Table 8** below

- H₀: Participants would react to political news stories without regard to their level of political group identification.
- H₁: Participants with higher political group identification would be more willing to share fake news about the other group.
- H₂: Participants with higher political group identification would be more willing to flag genuine news about their own group as fake.

H₀, the null hypothesis, represents the baseline against which the other hypotheses are tested. It assumes the case where no relationship appears between the participants' level of group identification with one or the other political party and whether they would choose to share a political news story or flag it as fake. The test hypotheses assume the cases where stronger political group identification is positively associated with the relevant reactions to political news.

Research Design

To make sure my research design produced reliable, valid results, I observed the following procedural considerations. First, I would need to collect enough information about my participants to give a good description of their overall characteristics. Study participants' social identity, in terms of political group-identification, would need to be assessed in a way that would be comparable between all participants, regardless of their political affiliation coming into the test. I also considered that the scope of the political group-identification assessment could be broadened or focused depending on the number of political affiliations I included in the evaluation. Testing my hypotheses would necessarily require news stories for the participants to encounter. Test items would need to be developed and presented in a way that could approximate news content on social media without introducing factors for which the test instrument did not or could not account. At the same time, the study would need to capture reactions to the news that reflect how users actually react on the platforms.

I had planned to survey participants from political parties in both the U.S. and U.K., taking advantage of the opportunity to test my hypotheses in proximity to the national elections in each country. I expected such timing would facilitate drawing in participants, and that testing in proximity to an actual election would also best interface with the dilemma often raised by many social media stakeholders: the potential for fake news to disrupt the

democratic process.⁷⁰⁶ Non-election years could provide opportunities for comparative testing to see how the proximity of an election influences political group-identification and whether any such effect might apply generally across sample populations, as upcoming elections tend to make political identity more salient.⁷⁰⁷ Given that the U.S. conducts national Congressional elections every two years and Presidential elections every four years, the predictable timing of the American process in particular allows for the routine replication of my experiment. Nevertheless, the further away in time from the present study that any subsequent test may be conducted, the more likely external conditions around the experience of political news on social media might change.⁷⁰⁸

My original plans regarding participant recruitment and assessment had to be modified when the U.K. General Election was called early,⁷⁰⁹ taking place in December 2019 before my experiment was ready to launch. Despite this, the U.S.-only case presented an opportunity to limit the scope of the assessment to only two political affiliations,⁷¹⁰ as just two political parties dominate U.S. politics and any other substantial political movements there can be taken largely as extensions of the major parties' institutions and ideologies.⁷¹¹ Future U.K. elections might provide the opportunity to recruit a completely distinct sample population for comparison and also to assess different political affiliations that do not merely mirror the U.S. political spectrum—for example, between unionist and independence parties in Scotland or Northern Ireland.

I also wanted to ensure the test would translate broadly to the experience of news content on social media while maintaining close control over what the study participants were reacting to specifically. An unmanaged design aspect can introduce a 'confound' into the experiment where it invites responses from participants that reflect something other than what the researcher intends to measure. For example, the plausibility of a hypothetical dilemma could become an overriding concern in the evaluation of moral judgments, rather

⁷⁰⁶ Conceivably, this experiment could be constructed around participants from any two groups for whom news stories on social media could reliably implicate their social identity (such as rival football teams), but studying political group members best supports the broader aims of my Thesis

⁷⁰⁷ Discussed in Part III, 'Principles of Social Psychology' – Social Identity Theory

⁷⁰⁸ *e.g.*, the social media platforms make substantial changes to their content policies and moderation procedures between test runs

⁷⁰⁹ UK Parliament, 'General Elections' (2021) <<https://www.parliament.uk/about/how/elections-and-voting/general/>> accessed 28 April 2021

⁷¹⁰ A trans-Atlantic approach might have framed the groups as 'conservative' and 'progressive', although those descriptors are more nebulous than party affiliation.

⁷¹¹ Pew Research Center, 'Political Independents: Who They Are, What They Think' (14 March 2019) <<https://www.pewresearch.org/politics/2019/03/14/political-independents-who-they-are-what-they-think/>> accessed 28 April 2021 (reporting that 81% of independent voters "lean" toward one of the two major parties)

than the positive or negative moral valence of the dilemma.⁷¹² Making the test closer to the real-world experience of social media users would make for a more direct comparison but could also introduce more factors that could mask or distort what I wanted to observe.⁷¹³ On the other hand, abstracting further away from the messy real-world environment would reduce the interference of other factors but could also make the test less representative of reality.

I initially considered a design in which the news stories would be presented as graphical mock-ups of actual social media sites' news posts. Following a presentation I gave at UEA while still developing this 'screenshot' approach, an attendee remarked that he thought he knew that a particular story I had shown was true but had nevertheless labelled it 'fake' because it appeared to come from tabloid newspaper *The Sun*. That suggested to me that including information about the source of a story would invite participants to make judgments about the source itself rather than the subject matter in ways that could be difficult to screen out when looking at the test results. Excluding or occluding images, links, logos, and other identifiers sufficiently enough to remove their influence from the participants' reactions to the news items while maintaining that level of verisimilitude likely would have proven impracticable for this study. Future studies aimed at building on research into media modality effects⁷¹⁴ might look to how additional graphical information (such as a photograph accompanying the story, or a user's avatar or profile picture) or the story's appearing on one particular social media platform or another might affect decision-making about content.⁷¹⁵

Aiming to have the participants base their reactions on the mental pictures, feelings, and values they associate with the groups indicated in the stories, I opted for a design that minimized the amount of prejudicial information participants would be able to pull from the test items. My approach was to provide the level of informational context people might see scrolling past a post on social media: a headline and quick blurb comparable to the amount of textual information in a tweet or News Feed caption concerning a typical public affairs news

⁷¹² Anita Körner, Susann Joffe, and Roland Deutsch, 'When skeptical, stick with the norm: Low dilemma plausibility increases deontological moral judgments' (2019) 84 J Experimental Social Psychology 103834

⁷¹³ For example, if I provided subjects with full-length news articles to read, other measures would be necessary to control for exogenous factors that are not present (or else managed) in the limited context of the headline-and-blurb approach, and fewer test items would be feasible to include.

⁷¹⁴ *i.e.*, something updating the study in Shelly Chaiken and Alice H Eagly, 'Communication Modality as a Determinant of Persuasion: The Role of Communicator Salience' (1983) 45 (2) J of Personality and Social Psychology 241

⁷¹⁵ Holly Garnett, Royal Military College of Canada, has been developing a study along these lines, comparing user experiences across social media platforms to investigate how differences between platforms (such as open/closed network structure, level of interactivity) might contribute to the believability of genuine or fake news sources; Holly Garnett, 'Presentation' (Politics on Wednesdays, UEA School of Politics, Philosophy, and Language, 28 September 2022)

event. I thus expected participants' reactions to the test items would reflect comparable contributions by the same mental processes at work in similarly brief real-world experiences of political news on social media. In order to engage the political aspect of the participants' social identities, my test items would implicate the partisan groups relevant to my participants by indicating the party affiliation of elected officials involved in matters of public concern.

Additionally, I recognized that the number of items I included would mean still other trade-offs and could affect response quality and completion rates.⁷¹⁶ More items could generate more data, but too many test items could wear down the attention span of the participants or introduce other effects as the reaction 'task' went along. I decided to pilot test the experiment design amongst a selection of volunteers. Obtaining feedback from the pilot participants on how long the experiment took them to complete and what their subjective impressions of it were helped me make sure that, in the experiences of my final participants, the most important differences between the test items would be their relevance to the participants' political group-identities. Doing so allowed me to include a substantial number of test items while keeping the duration of the study manageable and completion rates high.

I also reasoned that eliciting relevant responses from my participants would require careful consideration of the tone of the test items. Whether the stories had a positive, neutral, or negative valence for the participants could influence how they reacted as well as the strength of those responses.⁷¹⁷ However, testing all those valence conditions in concert could introduce a number of substantial challenges: assessing whether reactions to the positive, neutral, and negative stories would be directly comparable, and whether the combined presence of those test conditions itself might introduce changes in my participants' decision making that would be difficult to discriminate statistically.⁷¹⁸ As discussed in Part Three, the political campaign study by Bizer, Larsen, and Petty found that negatively framed attitudes demonstrated "higher levels of attitude certainty and attitude-consistent behavioral

⁷¹⁶ Though always planning to compensate my participants for their time, I also recognized that I might not be able to rely on much beyond "market rates" to predict the point at which they might decide that the careful completion of my study was no longer worth it—compensation is discussed further in the Ethics section below

⁷¹⁷ *e.g.*, George Y Bizer, Jeff T Larsen, and Richard E Petty, 'Exploring the Valence-Framing Effect: Negative Framing Enhances Attitude Strength' (2011) 32 (1) Political Psychology 59 (uncovering differences in participants' reactions to "preference prompts" in political contexts due to framing manipulation where a prompt is positively framed with regard to the subject's support for one candidate or negatively framed as opposition to another)

⁷¹⁸ *Compare* Fiona Grant and Michael A Hogg, 'Self-uncertainty, social identity prominence and group identification' (2012) 38 J Experimental Social Psychology 538 *with* Joseph A Vitriol and others, 'Ideological uncertainty and investment of the self in politics' (2019) 82 J Experimental Social Psychology 85 (suggesting competing pressures in the political identification context might lead to ambiguous reactions)

intentions.”⁷¹⁹ This suggested that I would get clearer reaction signals from my participants where they rejected rather than affirmed news about their political party. I therefore decided I would test only for reactions to stories with negative valence for the political groups under assessment.

By using all negative stories, the design aimed to ensure participants’ reactions to the different test items would be more directly comparable to one another and would not be influenced by effects arising incidental to the participants’ contemporaneous interaction with positive stories. Additionally, the reactions of one affiliation of study participants to news stories about the other could be evaluated in light of those participants’ reactions to news stories critical of their own political group, and the same pool of stories could be used for repeating the evaluation of those reactions from the other side. In keeping with the goal of developing a valid but realistic testbed, I identified the following story parameters as criteria for balancing the subject matter of my test items: notoriety of the person(s) involved, i.e. level of political influence or representativeness; geographic distribution; political issues or values at stake; outrage or salaciousness; and word choice. The topics of the news items included the following matters of public concern: insider trading, campaign financing, nepotism, public corruption, fraud, and sexual misconduct or other malfeasance by public officials.

The final (but perhaps most consequential) aspect of test item design which I considered was how to reliably determine what would constitute a true or fake news story. I recognized that a fundamental ontological question had to be addressed first: What makes a news story true? Here again the insights of Lippmann are invaluable, particularly where he sets out the distinction between news and truth:

“The function of news is to signalize an event, the function of truth is to bring to light the hidden facts, to set them into relation with each other, and make a picture of reality on which men can act. Only at those points, where social conditions take recognizable and measurable shape, do the body of truth and the body of news coincide. . . . In this sector, and only in this sector, the tests of the news are sufficiently exact to make the charges of perversion or suppression more than a partisan judgment.”⁷²⁰

This proposition of Lippmann’s interested me particularly because my experiment would seek to characterize and quantify the extent of that very same “partisan judgment” he

⁷¹⁹ Bizer, Larsen, and Petty (n 717); see also Part III, ‘The Social Psychology of Political News and Social Media’

⁷²⁰ Walter Lippmann, *Public Opinion* (Simon & Schuster 1997) 226

referenced. I recognized accordingly that the ‘truth’ of news items I included in my experiment would need to be assessed in relation to whether they represented actionable information about the matters asserted in the stories. In other words, the genuine news stories I included would need more than their mere publication in major circulation media sources. These news stories would need to represent verifiable events⁷²¹ in the space of public life which had measurable consequences for the political affiliations involved—such as the prosecution of one of a party’s elected officials, or the public statement of a policy position by one of a party’s candidates for office.

Key inferences from my research discussed in previous Parts suggested that both lay-person and expert declarations of ‘fake news’ may not be especially reliable, particularly in the context of public affairs news on social media, where the accusation may serve political or commercial ends as well as engage group-related psychological processes of identification and self-esteem management.⁷²² I therefore decided my design would avoid relying on the presentation of news stories which had been labelled false or fake by media watchdogs, news retailers, or other stakeholders. Sticking with Lippmann’s proposition, I considered that so long as a story signified an event, it really would be ‘news’—whereas many of the aforementioned media critics are sometimes concerned with purported ‘news’ that has signalled a *non-event* such as an opinion (meaning ultimately that it isn’t ‘fake’ news, but rather that it is not ‘news’ at all, sticking with Lippmann’s distinction)⁷²³ or encoded a *genuine* event in a way that sends the ‘wrong’ signal to supposedly credulous audiences.⁷²⁴

Accordingly, I decided that I could best test my hypothesis by making up fake *news* stories out of whole cloth: stories that would realistically signify completely fictitious events involving imaginary people with no connection to the study participants’ frames of reference other than textually indicated political affiliation, geography (due to the nature of U.S. politics), and plausible circumstances suggestive of public action like a court case or electoral vote.⁷²⁵ As a result, my participants would face situations where the possibility of a news story being fake (or not) might be overcome by the participants’ social identification-related motivations in deciding to flag or share the story. Likewise, participants would stand a lower chance of coming into the study with sufficient knowledge to reject a story as false based

⁷²¹ i.e., events for which evidence of their existence could conceivably be obtained and tested by an outside party

⁷²² Part I, ‘Fake News Studies’; Part II, ‘Platform Policies and Technical Capabilities’ – Evaluating Content Moderation at Scale; Part III, ‘The Social Psychology of Political News and Social Media’

⁷²³ Lippmann (n 720) at 226-28

⁷²⁴ *ibid* at 142-44, 153

⁷²⁵ *ibid* at 142-43 (Lippmann points out news readers’ reliance on authoritative or “trustworthy” reporters when faced with the challenge of distinguishing true and false stories of such characteristics—as discussed above, I had determined my experiment would exclude such sourcing information to remove that very factor from participants’ decision making)

solely on prior awareness of the facts of a particular event. This would achieve my design objective while avoiding the aforementioned media studies pitfalls.

Test Procedure: Measuring Social Identification

A major concern for psychological research designers is to make sure they are actually getting the results they aim to collect. Data can be skewed by a measurement itself or a respondent's tendency to respond in a self-affirming way,⁷²⁶ for example. Likewise, choosing the right research tools in social psychology is especially important because of how pervasive some of these mental processes are to how human interface with reality. Several scales have been developed over the years to reliably measure the strength of an individual's subjective relationship with their perceived group.⁷²⁷ In their study of work-organization self-categorization and commitment, Bergami and Bagozzi developed a novel assessment of identification, using a Venn diagram-style visual item and verbal scale.⁷²⁸ Their investigation looked into the content of participants' stereotypes, and they conducted preliminary interviews to populate their measures of participants' value connotations of their work organization rather than assume appropriate framing *a priori*.⁷²⁹ Other attempts have also been made to develop reliable, parsimonious assessments of social identification in even as brief as a single-item measurement.⁷³⁰ Fortunately, a widely accepted and relatively practical approach is to ask people to indicate on a scale how they see themselves.⁷³¹

For my study, I needed to learn a key piece of information about each of my participants: whether they identified more or less strongly with Republicans or Democrats. With this information in hand, I could then examine the relationship between that characteristic of my participants and their reactions to political news stories such as they might see on social media. Accordingly, I asked my participants a series of 'self-report' questions related to their political identity. Participants were asked about their voter registration status and given the option to choose yes, no, or not sure. Participants were then asked if they considered themselves a Republican or Democrat, the two major parties in

⁷²⁶ Anton J Nederhof, 'Methods of Coping with Social Desirability Bias: A Review' (1985) 15 *Eur J Social Psychology* 263, 275-76 (while no specific method may completely address this "social desirability bias" in experimental and survey research, combined prevention and detection methods help reduce the effect)

⁷²⁷ Massimo Bergami and Richard P Bagozzi, 'Self-categorization, affective commitment and group self-esteem as distinct aspects of social identity in the organization' (2000) 39 *British J Social Psychology* 555, 555-64

⁷²⁸ *ibid* 564-65

⁷²⁹ *ibid* 565

⁷³⁰ Tom Postmes, S Alexander Haslam, and Lise Jans, 'A single-item measure of social identification: Reliability, validity, and utility' (2013) 52 *British J Social Psychology* 597

⁷³¹ Delroy L Paulhus and Simine Vazire, 'The Self-Report Method' in Richard W Robbins, R Chris Fraley, and Robert F Krueger (eds), *Handbook of Research Methods in Personality Psychology* (Guilford Press 2007) 224

the U.S.⁷³² Because not all participants will consider themselves one or the other,⁷³³ a subsequent ‘forced choice’ question asked them to choose which of the two parties is closer to their own politics.⁷³⁴ This inflection point provided my study an additional advantage: participants can be sorted into groups of self-labelled Republicans and Democrats and of those who do not describe themselves in those terms but still be evaluated in terms of those two groups. The complete test instrument is included in **Appendix A**.

Based on their responses to the forced choice question, participants were then directed to a set of questions specific to one or the other political affiliation. I adapted these questions from the four item group identification scale used by Smith, Seger, and Mackie.⁷³⁵ Participants who chose ‘Republican’ as closer to their own politics received a four item assessment of their Republican identification, while participants who chose ‘Democrat’ received a four item assessment of their Democratic identification. Other than the affiliation identifier (the name of the party), the items were identical for each group. Participants were directed to rate each of the four items on a seven point Likert-type scale anchored at point one by ‘do not agree at all’ and at point seven by ‘agree completely.’ Based on their ratings of these items, participants were assigned a score that reflects their average level of political group-identification. Using the identification scale in this way accommodates participants who might otherwise be awkwardly shoved into either group, with those who previously self-described as Republican or Democrat likely indicating a higher identification score than those who did not, which in turn helps the data analysis by providing a broader range of scores.

Test Procedure: Presenting News Content Items

The next section of the survey featured the real and fake news items participants were asked to evaluate. The instructions introducing this section read as follows:

The following news stories come from a popular social media site. Please read each item carefully and respond to the accompanying prompts just by giving your first impressions of what you read.

⁷³² A slightly different set of questions might be deployed for a U.K. test to better reflect the U.K.’s formalized party membership system, such as first asking participants if they usually vote for Labour or the Conservative Party, and then ask whether they are registered members of a political party.

⁷³³ Notably, many American “leaners” who identify as independent voters tend nevertheless to express fairly consistent party voting preferences; see, *e.g.*, John R Petrocik, ‘Measuring party support: Leaners are not independents’ (2009) 28(4) Electoral Studies 562

⁷³⁴ Eliot R Smith, Charles R Seger, & Diane M Mackie, ‘Can Emotions Be Truly Group Level? Evidence Regarding Four Conceptual Criteria’ (2007) 93(3) J Personality and Social Psychology 431, 434

⁷³⁵ *ibid*

The instructions were provided in this fashion to promote several objectives. First, the instructions needed to indicate the social media context in which such items could be encountered in the wild. Second, the instructions needed to encourage participants to read and reflect on the stories sufficiently to make a judgment about them, but not suggest participants needed to do any external investigating in search of a ‘correct’ answer to the prompts—thus only first impressions were requested. Third, no indication of the items’ possible veracity or falsity was given in the instructions to avoid predisposing the participants one way or the other.

The stories themselves were presented visually as bolded text headlines and normal typeface brief introductory sentences (or, “ledes”) describing the subject matter of each story. An example item follows:

Peer pressure amidst Congress members over stock trades

U.S. Representative Bill Gulch, an Iowa Republican, is under investigation by the House Committee on Ethics after he sold Monsanto stock ahead of the public release of findings from a closed hearing.

Twenty headline-and-lede news items were included in the final design. To reduce the likelihood that any testing effects would arise over the course of the participants’ exposure to the collection of news items, the number of stories involving Democrats matched the number of stories about Republicans and the order in which the stories were presented was randomized for each participant. While the actual truth or falsity of the stories was not made apparent to the participants, the number of true stories likewise matched the number of false stories.

Test Procedure: Measuring Reactions to News Content

Robustly capturing participants’ reactions would require measuring their experience of the test items along dimensions that are meaningfully related to how the broader population interacts with social media news posts. At the same time, those measurements would need to provide reliable, comparable data. As discussed in Part Two, the spread of fake news or suppression of genuine content on social media can happen through various mechanisms which are developed and maintained by the platforms themselves: direct mechanisms like sharing or flagging by users and indirect mechanisms like promoting or down-ranking content through the platforms’ automated processes.⁷³⁶ A major challenge with any measure would be in estimating how platform algorithms might translate such

⁷³⁶ Part II, ‘Platform Policies and Technical Capabilities’

information into the moderation of a story. Instead, my goal here was to reveal any extent to which the fact of a user's political group identification might relate to the kinds and quality of data inputs the platform algorithms rely upon when incorporating user reactions to political news content.

To achieve the level of detail and understanding I wanted, I decided I would use Likert-type scale items⁷³⁷ to record participants' self-reported likelihood to share or flag each news item. Scale items would allow for comparison between participants' intensity of intention to share or report the news items from each test condition and obviate the need to score or rank the news items themselves as being more or less 'Democrat' or more or less 'fake'—something that might be necessary for determining identity association if a 'Share' button press was the only participant reaction data recorded. The Likert scale approach would also allow for an objective check of item satisfactoriness by measuring the internal consistency of the scale items.⁷³⁸ For example, if all of my participants, Republicans and Democrats alike, each indicated that a specific news story should definitely be reported as fake, such a result might signal that the news story was too obviously fake to provide a useful test or that the subject matter was preoccupying.

Participants were presented with a prompt after each item asking them to think about what they read and then respond on two seven point-Likert type scale items anchored at point one by "Definitely would not" and point seven by "Definitely would" with regard to whether they were willing to share the news story on social media or to flag or report the story as fake.⁷³⁹ Points two through six were labelled with their numerical value but assigned no further suggestion of interval between the two anchors. The scale items allowed participants to indicate a range of reactions that could be assigned a numerical value for later analysis and comparison. This approach also captures situations where social media users both share *and* flag a story (so as to draw attention to the outrageousness of a particular claim, for instance). While not the same as measuring actual user behaviours on social media platforms, measuring participant intentions like this serves as a useful proxy by targeting the corresponding decision-making involved in 'real world' behavioural performance, as similar experimental studies often presume.⁷⁴⁰ Furthermore, I had no *a priori* reason to believe my

⁷³⁷ Rensis Likert, 'A Technique for the Measurement of Attitudes' in R S Woodworth (ed) *Archives of Psychology* 22 (1932-33) 5, 26; 44-46

⁷³⁸ *ibid* 48-51

⁷³⁹ While not the central focus of my study, additional measurements capturing emotional reactions to each news item were also taken and are described in **Appendix C**.

⁷⁴⁰ See, e.g., Icek Ajzen and Martin Fishbein, 'Attitudes and the Attitude-Behavior Relation: Reasoned and Automatic Processes' (2000) 11 *European Review of Social Psychology* 1, 16-17 (discussing the evolution of studies predicting behaviour from attitudes, with regard to compatible contexts and criteria)

participants' actual flagging and sharing behaviours would be any less extreme than in my study. These prompts appeared as follows:

Thinking about what you read...

	Definitely would not 1	2	3	4	5	6	Definitely would 7
Would you share this news story on social media?	☐	☐	☐	☐	☐	☐	☐
Would you flag or report this story as fake on social media?	☐	☐	☐	☐	☐	☐	☐

Ethics

Ethics approval was sought through the UEA School of Psychology. I submitted my application, attached as **Appendix B**, through the School's online portal on September 9th, 2020. Considerations I reviewed included privacy concerns, appropriate safeguards for research over the Internet, and potential sensitivity of the topics presented in the test items. Participant privacy was assured through appropriate anonymization during the data collection process and later data retention procedures. Online research carries various perils related to verifying participant eligibility (such as age and location) as well as handling compensation, which were managed through the recruitment process discussed below.

Additional safeguards were undertaken through the design of the experiment itself (along with its pilot testing) to respect the value of participants' time. As for managing the news story topics, while some subjects like sexual harassment were potentially sensitive for some participants, all of the stories were presented in approximately the same fashion as someone might casually see shared over social media and with no additional details into any of the events referenced. I subsequently received approval from the Ethics committee, with minor revisions related to confirmation of the participant compensation rate and information provided in the debriefing.

The consent form and debriefing were adapted from the School of Psychology's usual forms. Copies of each are included as part of **Appendix A**. The consent form explained participation in the study, including participants' freedom to withdraw at any time, and provided contact information if participants had any questions. The debriefing explained how participants' responses would be used and provided a final opportunity for participants to say whether they wanted their responses included.⁷⁴¹ The debriefing also let participants know that some of the news items were fabricated and that they were free to investigate further if they wanted to do so.

Participants

Participant recruitment was handled through CloudResearch, an online participant recruitment service integrated with the Amazon Mechanical Turk ("MTurk") "crowdsourced" online labour market.⁷⁴² CloudResearch evaluates MTurk worker characteristics to provide participant panels to researchers.⁷⁴³ Using their tools, I was able to select a pool of participants exclusively located in the U.S. Compensation for participants was handled through MTurk—\$1.50 each. My study was advertised to MTurk workers as a psychology study, about fifteen to twenty minutes long, concerning participants' impressions of news stories on social media ahead of the 2020 U.S. elections. Online delivery of the research questionnaire was handled via the Qualtrics Platform.⁷⁴⁴ Qualtrics access was provided through the UEA School of Psychology. Qualtrics' features were used in particular to accomplish the political group-identification assessment and to randomize presentation of the test items, as well as to generate random ID numbers for participants.

The experiment was opened to participants on October 1st, 2020 at 2:00 PM UTC. Thirty-two participants were obtained before the study shut down prematurely due to an error in the way funding had been provisioned for participant compensation. I contacted Amazon's support staff and was eventually able to have my usage limit adjusted upwards. I reopened the study on October 6th, 2021 and drew in 374 additional participants. Completed studies from both runs were included in my final dataset. After my study closed, I removed unfinished and debrief-withdrawn surveys, leaving 396. I then took a look at the duration data and test item responses from the surveys to check for possible 'speeding' behaviours

⁷⁴¹ Rather than have an attention check item appear randomly during the study, the attention check item was included at the end (before the debrief) with the aim of having participants give due consideration to the level of attention they paid to the preceding questions before deciding on response inclusion

⁷⁴² CloudResearch (2021) <<https://www.cloudresearch.com>> last accessed 4 May 2021; Amazon Mechanical Turk (2018) <<https://www.mturk.com>> last accessed 4 May 2021

⁷⁴³ Prime Research Solutions, 'How Does CloudResearch know Panel Demographic Data?' (CloudResearch.com, 2020) <<https://go.cloudresearch.com/knowledge/how-does-turkprime-know-panel-demographic-data>> last accessed 4 May 2021

⁷⁴⁴ Qualtrics XM (2021) <<https://www.qualtrics.com/>> last accessed 4 May 2021

(where online survey participants simply click through the items rapidly to reach the end).⁷⁴⁵ The average survey duration was 813.4 seconds, and surveys completed in less than 416.444 seconds (just under 7 minutes) were more than one standard deviation below the mean completion time. 39 surveys fell below this mark. Upon closer scrutiny, many of these surveys exhibited a marked lack of answer variation, as would be expected in survey speeding. I removed the 39 ‘fast’ surveys from consideration, leaving 357 ‘quality’ surveys for analysis.

A power analysis⁷⁴⁶ conducted using G*Power indicated that my study sample would be sufficiently sensitive (power = .80) to detect significant differences ($p < .05$) between my Republican participants ($n = 139$) and Democrat participants ($n = 218$) at or above a small-to-moderate effect size (Cohen’s $d = .328$).⁷⁴⁷ Additional analysis showed that the sizes of my participant groups in turn would provide sufficient sensitivity (power = .80) to detect significant differences ($p < .05$) within each group at or above small effect sizes for both Republicans (Cohen’s $d = .258$) and Democrats (Cohen’s $d = .205$).⁷⁴⁸

The initial series of questions in the test instrument collected demographic information about the participants. Participants were asked to provide their age by entering an integer value in a blank field. They were next asked to provide information about the annual gross income of themselves and any family who lives with them. This question was constructed around the 2018 median household income amount in the U.S.,⁷⁴⁹ and participants were instructed to pick between less than \$62,000 and equal to or greater than that amount. Participants were then asked to provide information about their sex and given the options male, female, and other. Participants were also asked to describe their race or ethnicity and based on a list of choices, and participants who selected “Other” were provided a field in which to specify their responses. Geographical information about the participants was obtained via a drop-down menu listing each of the 50 U.S. States along with the District of Columbia, Puerto Rico, and several other Territories. Collecting this background demographic information not only describes who participated in the study but also facilitates checking whether variables other than political group-identification may be influencing

⁷⁴⁵ Chan Zhang and Fred Conrad, ‘Speeding in Web Surveys: The tendency to answer very fast and its association with straightlining’ (2014) 8(2) Survey Research Methods 127

⁷⁴⁶ See generally Jacob Cohen, *Statistical Power Analysis for the Behavioral Sciences* (Routledge 1988)

⁷⁴⁷ i.e., using a Mann-Whitney test to compare data values between groups, as described *infra*; reliably detecting differences at smaller effect sizes could potentially be achieved with larger sample sizes

⁷⁴⁸ i.e., using a Wilcoxon Signed Rank test to compare distributions between variables, as described *infra*

⁷⁴⁹ Gloria G Guzman, ‘Household Income: 2018’ (U.S. Census Bureau, Sept. 2019) American Community Survey Briefs ACSBR/18-01, 2

participants' responses to the news story test items.⁷⁵⁰ The demographics of the survey respondents are presented in **Table 1** below.

The minimum age of the participants was 18, and the maximum age was 69, with a mean of 39.57 years of age. 59.4% of participants reported they earned less household income than the national median household income for 2018, reflecting slightly more lower-income participants than expected.⁷⁵¹ Sex differences were as-expected, in line with U.S. Census Bureau numbers.⁷⁵² Participants' racial and ethnic characteristics were mostly in line with Census Bureau estimates as of 2019, which assessed the general U.S. population at 76.3% white, 13.4% black, 5.9% Asian, etc.⁷⁵³ However, hispanic/Latin American participation in the study (5.9%) was lower than the estimated national population rate of 18.5%.⁷⁵⁴

Lower engagement with hispanic/Latin American participants (relative to their presence in the U.S. population) could be due, to some extent, to the fact that the study advertisement I posted through CloudResearch was only provided in English, as was the study itself. While my sample was appropriately geographically limited and proportionally spread throughout the American states, additional language options could be provided to potentially improve the demographics of the participant pool of a second experiment. State residence reports were roughly in line with my expectations based on Census data,⁷⁵⁵ with the most populous states (California, Texas, New York, and Florida) returning the most survey responses and the least populous states and the District of Columbia returning the fewest. No responses were returned from Alaska, Montana, New Hampshire, North Dakota, Utah, Puerto Rico or the U.S. Virgin Islands.

⁷⁵⁰ For example, checking whether household income responses show a stronger relation than party affiliation to responses to a news story about a profligate politician

⁷⁵¹ Guzman (n 749)

⁷⁵² U.S. Census Bureau, Current Population Survey: Annual Social and Economic Supplement 2019 (April 2020) <<https://www.census.gov/data/tables/2019/demo/age-and-sex-composition.html>> last accessed 05 October 2021

⁷⁵³ U.S. Census Bureau, 'QuickFacts: United States' (July 2019)

<<https://www.census.gov/quickfacts/fact/table/US/PST045219>> last accessed 05 October 2021

⁷⁵⁴ *ibid*

⁷⁵⁵ U.S. Census Bureau, 'Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2019' (April 2021) <<https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html>> last accessed 05 October 2021

Table 1*Participant Demographics*

Variable	Minimum	Maximum	Median	Mean	Std. Deviation
Age	18	69	37	39.57	11.569

Variable	Frequency	Percentage
Income		
<i>Less than \$62,000</i>	212	59.4
<i>\$62,000 or greater</i>	145	40.6
Sex		
<i>Male</i>	176	49.3
<i>Female</i>	179	50.1
<i>Other</i>	2	0.6
Race		
<i>White or European American</i>	261	73.1
<i>Hispanic, Latin or South American</i>	21	5.9
<i>Black or African American</i>	43	12.0
<i>American Indian or Alaskan Native</i>	1	0.3
<i>East Asian</i>	19	5.3
<i>South Asian</i>	8	2.2
<i>Other (please specify)</i>	4	1.1
State*		
<i>California</i>	36	10.1
<i>Texas</i>	33	9.2
<i>Florida</i>	30	8.4
<i>New York</i>	26	7.3
<i>Illinois</i>	15	4.2
<i>...</i>		
<i>South Dakota</i>	1	0.3
<i>Rhode Island</i>	1	0.3
<i>Vermont</i>	1	0.3
<i>West Virginia</i>	1	0.3
<i>Wyoming</i>	1	0.3

* Top five and bottom five shown here for sake of brevity, remainder available upon request.

Findings

Measuring Social Identification

Participants' responses to the four item-identification measures adapted from Smith, Seger, and Mackie⁷⁵⁶ provided the key indicator of group identification in my study.⁷⁵⁷ Responses to the four items showed good consistency with one another, indicating the reliability of the scale.⁷⁵⁸ The aggregated group identification scores for my participants were fairly high on average,⁷⁵⁹ and the distribution of the scores showed a significant difference from the standard normal distribution.⁷⁶⁰ Accordingly, I have used statistical tests for non-standard distributions to analyse my findings, described briefly in the footnote below.⁷⁶¹ Based on the forced choice question, I sorted my participants into Republican and Democrat groups. The groups showed no significant differences in terms of group identification scores,⁷⁶² indicating that the participants in one group were, on average, no more or less strongly identified with their particular group than participants in the other group.⁷⁶³ The two groups likewise showed no significant differences in terms of

⁷⁵⁶ Smith, Seger, and Mackie (n 734)

⁷⁵⁷ Additional measurements supporting construct validity are discussed in **Appendix C**

⁷⁵⁸ Cronbach's Alpha = 0.961 (a higher alpha statistic indicates greater internal consistency between items)

⁷⁵⁹ Mean = 5.2073, Standard Deviation = 1.56589

⁷⁶⁰ Kolmogorov-Smirnov D = 0.126, $p < 0.001$; skewness = -1.038; kurtosis = 0.582 (a significant D statistic in this test indicates that the distribution of scores differs from the standard normal distribution's bell shaped curve, where most of the results lie around the middle and outliers fall evenly on one side or the other)

⁷⁶¹ Non-parametric tests do not require data to be normally distributed; below are brief descriptions of the tests I used:

Mann-Whitney test: compares data values from two groups (such as my Republican and Democrat group identification scores); a significant Z statistic ($p < 0.05$) in the M-W test indicates a substantial difference in data values between groups

Kolmogorov-Smirnov test: compares how data (such as my state residence demographic) about two groups are distributed; a significant Z statistic ($p < 0.05$) in the K-S test indicates a substantial difference in distributions between groups

Wilcoxon Signed Ranks test: compares how pairs of ordered variables are distributed (such as my sharing/flagging scales for genuine and fake stories); a significant Z statistic ($p < 0.05$) in this test indicates a different distribution between the paired variables

Kendall Rank Correlation test: measures the strength and direction of correlation between two variables (such as my group identification scores and my sharing/flagging scales); a significant tau-beta statistic ($p < 0.05$) in this test indicates a substantial relationship between the variables, and the sign (+/-) of tau-beta indicates the direction of the relationship, whether positive or negative

⁷⁶² Mann-Whitney (Z = -1.096, $p = 0.273$)

⁷⁶³ This does not, however, indicate that the groups are the same or that no actual difference exists—just that differences (if any) within the online sample are quite small and relatively unimportant—and in any case, group identification is controlled for in the analyses which follow.

demographics.⁷⁶⁴ The high scores could however indicate some selection bias resulting from the timing of the study in proximity to the election—a second run of the experiment during the off-season could show if that created a strong effect. A comparison between the participant groups' identification scores is shown in **Table 2**.

Table 2

Comparison Between Participant Groups by Group Identification Scores

Republicans (n = 139, 38.9%)		Democrats (n = 218, 61.1%)	
Mean = 5.095 SD = 1.584		Mean = 5.279 SD = 1.553	
<i>Mean Rank</i>		<i>Mean Rank</i>	<i>Z-value</i> <i>p</i>
171.54		183.76	-1.096 .273

Measuring Reactions to News Content

Participants' reactions to the twenty news story items on the sharing scale and flagging scale provided the data I compared with their group identification scores. Both of the scales showed high reliability across the test item responses.⁷⁶⁵ Comparisons between the participant groups on both scales are shown in **Table 3**. The sharing scale results showed that on average, each participant group was as likely as the other to share the news stories they saw in the study.⁷⁶⁶ That held true whether the stories were genuine⁷⁶⁷ or fictitious.⁷⁶⁸ The flagging scale results likewise showed that on average, each participant group was as likely as the other to flag the news stories they saw in the study as false.⁷⁶⁹ That also held true whether the stories were actually genuine⁷⁷⁰ or fictitious.⁷⁷¹

⁷⁶⁴ Age ($Z = -1.875$, $p = 0.061$); Income ($Z = -0.321$, $p = 0.748$); Sex (Kolmogorov-Smirnov $Z = 0.486$, $p = 0.972$); Race (Kolmogorov-Smirnov $Z = 0.375$, $p = 0.999$); State Residence (Kolmogorov-Smirnov $Z = 0.715$, $p = 0.686$)

⁷⁶⁵ *Sharing* (Cronbach's Alpha = 0.972); *Flagging* (Cronbach's Alpha = 0.976)

⁷⁶⁶ Mann-Whitney ($Z = -1.026$, $p = 0.305$)

⁷⁶⁷ Mann-Whitney ($Z = -0.877$, $p = 0.380$)

⁷⁶⁸ Mann-Whitney ($Z = -1.153$, $p = 0.249$)

⁷⁶⁹ Mann-Whitney ($Z = -0.375$, $p = 0.708$)

⁷⁷⁰ Mann-Whitney ($Z = -0.694$, $p = 0.488$)

⁷⁷¹ Mann-Whitney ($Z = -0.145$, $p = 0.885$)

Table 3*Comparisons Between Participant Groups by Sharing and Flagging Scales*

	Republicans (n = 139)	Democrats (n = 218)		
	<i>Mean Rank</i>	<i>Mean Rank</i>	<i>Z-value</i>	<i>p</i>
<i>Sharing</i>				
All Stories	185.99	183.76	-1.026	.305
Genuine	184.97	175.19	-.877	.380
Fake	186.83	174.00	-1.153	.249
<i>Flagging</i>				
All Stories	181.52	177.39	-.375	.708
Genuine	183.63	176.05	-.694	.488
Fake	179.96	178.39	-.145	.885

Differences in sharing and flagging reactions based on news item category are shown in **Table 4** for Republican participants and **Table 5** for Democrat participants. Participants in each group were more willing to share stories about the other group than their own,⁷⁷² which was in line with my expectations given the negative valence of the stories.⁷⁷³ When participants in either group indicated they were likely to share stories that were actually fakes, they did so for the fake stories about the other side more than their own.⁷⁷⁴ Given that participants had no more reason to know a made-up story about one side was actually fake than a made-up story about the other (but in any case knew which party was involved), this suggests some motivation other than desire for accuracy or fairness. Participants in each group also tended to flag stories about their own party more than about the other,⁷⁷⁵ as I expected based on the social identity theory research discussed in Part Three.⁷⁷⁶ Likewise,

⁷⁷² *Republicans*, Wilcoxon Signed Ranks ($Z = -3.693$, $p < 0.001$); *Democrats*, Wilcoxon Signed Ranks ($Z = -6.375$, $p < 0.001$)

⁷⁷³ Discussed *supra*, 'Methodology' – Research Design; Part III, 'The Social Psychology of Political News and Social Media'

⁷⁷⁴ *Republicans*, Wilcoxon Signed Ranks ($Z = -4.373$, $p < 0.001$); *Democrats*, Wilcoxon Signed Ranks ($Z = -5.089$, $p < 0.001$)

⁷⁷⁵ *Republicans*, Wilcoxon Signed Ranks ($Z = -3.682$, $p < 0.001$); *Democrats*, Wilcoxon Signed Ranks ($Z = -4.965$, $p < 0.001$)

⁷⁷⁶ Part III, 'Principles of Social Psychology' – Social Identity Theory

when participants in either group indicated they were likely to flag genuine stories as fake, they did so for genuine stories about their own side more than the other.⁷⁷⁷

Table 4

Republican Participant Sharing and Flagging Comparison by News Item Category

	Republican Stories	Democrat Stories		
	Mean Rank	Mean Rank	Z-value	p
<i>Sharing</i>				
All Stories	37.53	55.77	-3.693	<.001
Fake	28.48	52.59	-4.373	<.001
<i>Flagging</i>				
All Stories	46.39	33.38	-3.682	<.001
Genuine	40.06	32.71	-3.917	<.001

Table 5

Democrat Participant Sharing and Flagging Comparison by News Item Category

	Republican Stories	Democrat Stories		
	Mean Rank	Mean Rank	Z-value	p
<i>Sharing</i>				
All Stories	93.97	55.71	-6.375	<.001
Fake	81.43	58.67	-5.089	<.001
<i>Flagging</i>				
All Stories	53.51	69.98	-4.965	<.001
Genuine	52.60	62.52	-2.555	.011

The tendencies to share fake outgroup stories and flag genuine ingroup stories as fake represent both the two conditions relevant to my hypotheses and, in broader terms of content moderation and social media regulation, consequential systems inputs—particularly in the case of false positives.⁷⁷⁸ When I compared these tendencies between my participant

⁷⁷⁷ *Republicans*, Wilcoxon Signed Ranks ($Z = -3.917$, $p < 0.001$); *Democrats*, Wilcoxon Signed Ranks ($Z = -2.555$, $p = 0.011$)

⁷⁷⁸ Discussed further *infra*, ‘Implications’

groups, I found no significant differences between Republicans and Democrats.⁷⁷⁹ That comparison is shown in **Table 6**. To check for the likelihood of an erroneous conclusion of no statistical difference between the groups on these measures, I performed post hoc power analyses using G*Power. The observed power for the comparison of sharing fake outgroup stories was 0.59, and the observed power for the comparison of flagging genuine ingroup stories was 0.83. To achieve an actual power of 0.80 and detect a significant difference ($p < 0.05$) between groups in sharing fake outgroup stories at the same effect size (Cohen's $d = 0.195$) would have required a total sample size of 958, evenly divided between Republicans and Democrats.

Table 6

Comparison Between Groups by Relevant Decision Scores

Political Group			
Republicans (n = 139)	Democrats (n = 218)		
Mean Rank	Mean Rank	Z-value	p
<i>Sharing Fake Outgroup Stories</i>			
188.74	172.79	-1.438	.150
<i>Flagging Genuine Ingroup Stories</i>			
184.44	175.53	-.821	.411

Analysis

To test for the hypothesized positive associations between political group identification and the observed sharing and flagging decisions, I conducted Kendall rank correlations for the relevant measures.⁷⁸⁰ Those correlations are shown in **Table 7**. I found a positive association between group identification scores and sharing scores for fake outgroup stories,⁷⁸¹ providing support for my first test hypothesis.⁷⁸² I also found a positive association between group identification scores and flagging scores for genuine ingroup stories,⁷⁸³ providing support for my second test hypothesis.⁷⁸⁴ Of further note, sharing fake

⁷⁷⁹ *Sharing*, Mann-Whitney ($Z = -1.438$, $p = .150$); *Flagging*, Mann-Whitney ($Z = -.821$, $p = .411$)

⁷⁸⁰ Additional correlational analyses of demographics and other measures is included in **Appendix C**

⁷⁸¹ Kendall Rank Correlation (tau-beta = 0.140, $p < 0.01$)

⁷⁸² Participants with higher political group identification would be more likely to share fake news about the other group.

⁷⁸³ Kendall Rank Correlation (tau-beta = 0.130, $p < 0.01$)

outgroup stories was also positively associated with flagging genuine ingroup stories.⁷⁸⁵ Each of these associations supports the notion that the study measurements are conceptually consistent with social identity theory. Group identification represents participants’ sense of similarity and feelings of satisfaction derived from affinity with a particular political group. Measurements of participants’ willingness to share fake stories that denigrate the outgroup or flag genuine stories that denigrate the ingroup *in relation to participants’ group identification* may thus highlight the social identity drives for group distinctiveness and positivity. The association between sharing and flagging in turn may reflect the relatedness of those drives to one another in constituting social identity.

Table 7

Kendall Rank Correlation Table

	Group Identification Score	Sharing Fake Outgroup Stories	Flagging Genuine Ingroup Stories
Group Identification Score	—		
Sharing Fake Outgroup Stories	.140**	—	
Flagging Genuine Ingroup Stories	.130**	.458**	—

Note. * p < .05, ** p < .01

I next expanded my correlation analysis to highlight group differences in the association between group identification scores and outgroup fake story share scores (as well as ingroup genuine story flagging). The results of these correlational analyses are shown below in **Table 8**. In the group separated analyses, only Republican participants showed a significant association between group identification and sharing fake outgroup stories. The absence of significant association among Democrat participants is discussed further below. Both groups, however, showed significant associations between the identification and flagging scores.

⁷⁸⁴ Participants with higher political group identification would be more likely to flag genuine news about their own group as fake

⁷⁸⁵ Kendall Rank Correlation (tau-beta = 0.458, p < 0.01)

Table 8*Group Separated Kendall Rank Correlation Table*

	Republican Group (n = 139)			Democrat Group (n = 218)		
	Group Identification Score	Sharing Fake Outgroup Stories	Flagging Genuine Ingroup Stories	Group Identification Score	Sharing Fake Outgroup Stories	Flagging Genuine Ingroup Stories
Group Identification Score	—			—		
Sharing Fake Outgroup Stories	.244**	—		.074	—	
Flagging Genuine Ingroup Stories	.151*	.537**	—	.124*	.397**	—

Note. * $p < .05$, ** $p < .01$

I then conducted regression analyses using SPSS to examine the extent to which group identification explained variability in these measures and to highlight any discernible differences in explanation between my groups.⁷⁸⁶ With regard to sharing fake outgroup stories, participant age (centered), sex/gender, and party affiliation were entered in Step 1, and average group identification score (centered) was added in Step 2.⁷⁸⁷ To evaluate the role of party affiliation on group identification's relationship to the sharing measure, an interaction term was created by multiplying the two variables and added in Step 3. The result of this regression is shown in **Table 9**.

⁷⁸⁶ Regression analysis does not perfectly fit my data and is undertaken here as a guide for further investigation and must be considered accordingly: for example, bias in the coefficients and p-values may be present due to differences in the variances between low and high group identity scores (e.g., someone with a low score will tend to have a low sharing measure, but someone with a high score might have high *and* low sharing measures)—I also performed weighted regression analyses, detailed in **Appendix C**, illustrating some of these phenomena, and explore how such analyses may be improved in future research below

⁷⁸⁷ Household income and voter registration status did not correlate with sharing or flagging and were excluded from the analysis; for additional details, see **Appendix C**

Table 9

Regression analysis predicting willingness to share fake outgroup stories from demographic factors and group identification scores, interaction

Variable	β	t	sr^2	F	R^2	ΔR^2
Step 1				9.253***	.073	—
Age	-.117	-2.260*				
Sex/Gender	-.214	-4.142***				
Party Affiliation	-.093	-1.810				
Step 2				12.967***	.128	.056***
Age	-.112	-2.231*	.012†			
Sex/Gender	-.241	-4.783***	.057			
Party Affiliation	-.105	-2.091*	.011†			
Group Identification	.238	4.736***	.056			
Step 3				12.349***	.137	.021**
Age	-.120	-2.405*	.014†			
Sex/Gender	-.240	-4.821***	.056			
Party Affiliation	-.109	-2.201*	.012†			
Group Identification	.715	4.232***	.043			
Party x Identification	-.499	-2.956**	.021††			

Note. * $p < .05$; ** $p < .01$; *** $p < .001$; †observed power $< .80$; ††observed power = .788

As expected, group identification explained a significant amount of the variance in participants' willingness to share fake outgroup stories: participants with higher group identification scores were more willing to share fake news about the other political side. Unexpectedly, age and gender both showed significant relationships to outgroup sharing, although the effect of age was below the sensitivity threshold of my study to characterize reliably.⁷⁸⁸ The negative coefficient of the interaction term, however, indicated that *which* party participants identified with may also explain part of that variance.⁷⁸⁹ To better interpret the interaction, I conducted separate regressions for each participant group, as shown in **Table 10** below. In the separated regressions, participant age (centered) and sex/gender were entered in Step 1, and average group identification score (centered) was added in Step 2.

⁷⁸⁸ I revisit these phenomena below in my discussion section

⁷⁸⁹ The effect size (Cohen's $f^2 = .021$) was slightly too small to be reliably captured at acceptable power (.80) given my sample size—368 total participants would have been sufficient show a significant difference ($p < .05$) at the same effect size

Among my Republican participants, group identification continued to explain a significant amount of the variance in sharing fake outgroup stories. The coefficients also indicated a significant negative relationship with age (and potentially sex/gender) among Republican participants, but the effect size was too small for reliable determination given my group sample size.

Table 10

Group separated regression analyses predicting willingness to share fake outgroup stories from demographic factors and group identification scores

<i>Republican Participants (n = 139)</i>						
Variable	β	t	sr^2	F	R^2	ΔR^2
Step 1				3.168*	.045	—
Age	-.172	-2.044*				
Sex/Gender	-.107	-1.265				
Step 2				11.200***	.199	.155***
Age	-.189	-2.437*	.035†			
Sex/Gender	-.152	-1.949	.023†			
Group Identification	.397	5.108***	.154			
<i>Democrat Participants (n = 218)</i>						
Variable	β	t	sr^2	F	R^2	ΔR^2
Step 1				11.034*	.093	—
Age	-.077	-1.179				
Sex/Gender	-.288	-4.409***				
Step 2				8.810***	.110	.017*
Age	-.069	-1.059	.005†			
Sex/Gender	-.303	-4.645***	.090			
Group Identification	.131	2.012*	.017†			

Note. * $p < .05$; ** $p < .01$; *** $p < .001$; †observed power $< .80$

Among my Democrat participants, group identification did not explain the variance in sharing fake outgroup stories to as great an extent,⁷⁹⁰ reflecting the significant interaction term returned in the preceding analysis. Group separated regression highlighted differences in the coefficients for group identification: the effect of group identification, with regard to

⁷⁹⁰ Cohen's $f^2 = .017$; observed power = .489

sharing fake outgroup stories, was much stronger among my Republican participants. Though smaller than that among Republicans, the group identification coefficient among Democrats still appears to move in the same direction.⁷⁹¹ Furthermore, among my Democrat participants, sex/gender showed a substantial explanation for variance in sharing fake outgroup stories, while age did not.

I then proceeded with a regression analysis regarding flagging genuine ingroup stories. The same demographic measures were again entered in Step 1, average group identification score (centered) was added in Step 2, and the interaction term between party and identification was added in Step 3. The result of that regression is shown in **Table 11**.

Table 11

Regression analysis predicting willingness to flag genuine ingroup stories as fake from demographic factors and group identification scores, interaction

Variable	β	t	sr^2	F	R^2	ΔR^2
Step 1				5.345***	.043	—
Age	-.116	-2.212*				
Sex/Gender	-.130	-2.488*				
Party Affiliation	-.104	-1.986*				
Step 2				10.067**	.103	.059***
Age	-.111	-2.183*	.012†			
Sex/Gender	-.159	-3.104**	.025			
Party Affiliation	-.116	-2.279*	.013†			
Group Identification	.245	4.819***	.059			
Step 3				8.356***	.106	.004
Age	-.114	-2.244*	.013†			
Sex/Gender	-.159	-3.100**	.024			
Party Affiliation	-.118	-2.315*	.014†			
Group Identification	.445	2.572*	.017†			
Party x Identification	-.209	-1.208	.003†			

Note. * $p < .05$; ** $p < .01$; *** $p < .001$; †observed power $< .80$

In this analysis, the addition of the interaction term did not return a statistically significant effect. That indicates insufficient evidence to support a conclusion that party

⁷⁹¹ Remaining positive within a 95% confidence interval for B; lower bound = .003, upper bound = .290

affiliation moderates the relationship of group identification with willingness to flag genuine ingroup stories as fake. With the interaction term excluded, group identification explained a significant amount of the variance in this ingroup flagging measure. As with the outgroup stories analyses, age and sex/gender both showed significant negative relationships with ingroup flagging, but the effect size of age remained below the sensitivity threshold of my study to characterize reliably.

To scrutinize the apparent absence of interaction between party and identification relative to the ingroup sharing measure, I conducted separate regressions for each participant group, as shown in **Table 12**. As previously, age (centered) and sex/gender were entered in Step 1, and average group identification score (centered) was added in Step 2.

Table 12

Group separated regression analyses predicting willingness to flag genuine ingroup stories as fake from demographic factors and group identification scores

<i>Republican Participants (n = 139)</i>						
Variable	β	t	sr^2	F	R^2	ΔR^2
Step 1				2.372	.034	—
Age	-.168	-1.983*				
Sex/Gender	-.060	-.705				
Step 2				5.460***	.108	.074***
Age	-.179	-2.196*	.032†			
Sex/Gender	-.091	-1.108	.008†			
Group Identification	.275	3.358***	.074			
<i>Democrat Participants (n = 218)</i>						
Variable	β	t	sr^2	F	R^2	ΔR^2
Step 1				5.144**	.046	—
Age	-.071	-1.065				
Sex/Gender	-.195	-2.906**				
Step 2				7.826***	.099	.053***
Age	-.057	-.869	.003†			
Sex/Gender	-.222	-3.378***	.048			
Group Identification	.233	3.554***	.053			

Note. * $p < .05$; ** $p < .01$; *** $p \leq .001$; †observed power $< .80$

Among both my participant groups, group identification explained a significant amount of the variance in flagging genuine ingroup stories as fake—in other words, higher group identification meaningfully corresponded with higher willingness to flag genuine stories about a participant's own group as fake, regardless of party. Coefficients for both groups showed a positive relationship between group identification and the ingroup flagging observations. Age showed a significant negative relationship among Republicans, but statistical power was again constrained by the small effect and sample sizes. Sex/gender continued to show a significant negative relationship with the ingroup flagging measure among Democrats but not Republicans.

Discussion

In this section, I have detailed the findings of my experiment to test the relationship between political group identification and reactions to social media news content. My correlational analysis showed a positive association between my study participants' level of identification as political group members and willingness to share fake news about rival political group members and to flag genuine news as fake when it concerns fellow political group members. My regression analyses suggested that this relationship may be to some degree attenuated by particular party affiliation in regard to sharing fake outgroup stories. However, considered in concert with the group separated correlational analyses showing no significant association between group identification and sharing fake outgroup stories among my Democrat participants, sharing fake outgroup stories may simply be less closely tied to the social identity motivations represented by group identification than I had originally assumed.

In terms of social identity theory, ingroup flagging appears to have a stronger conceptual 'fit' and clearer empirical support than outgroup sharing, given the regression results showing the positive relationship between group identification and flagging genuine ingroup stories regardless of party. Flagging negative true ingroup stories as fake may correspond more closely to maintaining ingroup positivity and distinctiveness⁷⁹² than does sharing fake stories that denigrate an outgroup—a more indirect means of satisfying those drives. Other accounts of motivated decision making may therefore potentially provide additionally robust explanations for observations of fake news sharing.

The difficulty of adequately characterizing the effect of group identification in the sharing test condition among my Democrat participants provide a useful frame for discussing several notable limitations of my study findings. First, obtaining a larger sample

⁷⁹² *i.e.*, as part of the esteem management 'toolbox' discussed in Part III, 'Principles of Social Psychology'

size may have allowed me to reliably discern the effect, assuming it would have remained the same among a larger pool of Democrats. Pursuant to adequate funding to pay more participants, a subsequent run of the experiment could be populated in better anticipation of smaller effect sizes.

Second, and more broadly in terms of experiment design, my study did not set out to investigate why Republicans and Democrats might differ in this regard—I had assumed they would not, based on the general principles of social identity as discussed in the preceding Part of this Thesis, but further investigation in this direction wasn't the aim of my study nor within the scope of my design. The difference in coefficients and effect size between my participant groups here could conceivably reflect an ideological asymmetry⁷⁹³ between the groups, wherein sharing fake outgroup stories uniquely goes against some part of the Democrat group 'doctrine'—but, as noted above, my Democrat participants did not significantly differ from my Republican participants in average sharing scores.⁷⁹⁴ Regardless, my study did not set out to investigate whether any specific ideological doctrines play a specific role in the decision to share fake news or not, *i.e.*, as a matter of party policy, beyond collecting measures of 'issue importance' concerning public discourse matters such as 'relying on facts' and 'hearing both sides' in addition to freedom of speech and freedom of the press. Detail on these measures and a brief assay into their relationship with the test conditions is included in **Appendix C**. Ultimately, further research may bring about a fuller picture of sharing fake news by following the structural framework suggested by Zmigrod⁷⁹⁵—paying sufficient heed to both doctrinal and relational aspects in the experiment design in order to appropriately illustrate potential effects and interactions. Such studies may prove especially relevant in crafting efficacious public policy responses to phenomena like fake news sharing.

Third in terms of limitations, and more broadly still, collecting my data in proximity to the 2020 U.S. Presidential election may have contributed to the skewness and kurtosis of my group identification measures—an exogenous factor pushing the measured values closer together and higher at the top end of the scale. That design choice was, as explained above,⁷⁹⁶ undertaken in part to capture conditions relevant to the public policy discussions of fake news interfering with elections central to this Thesis. Politics are nonetheless part of routine civic and social life, meaning that my observations here should not be taken as extraordinary or unusual. However, replicating the experiment in a non-election year could

⁷⁹³ Discussed in Part III, 'Principles of Social Psychology'

⁷⁹⁴ See **Table 6**

⁷⁹⁵ Leor Zmigrod, 'A Psychology of Ideology: Unpacking the Psychological Structure of Ideological Thinking' (2022) 17 (4) Perspectives on Psychological Science 1072

⁷⁹⁶ 'Methodology' – Research Design

potentially yield a group identification measure with a more normal distribution (facilitating the use of parametric tests) or with more stable variability in relation to test measures like outgroup sharing (improving confidence in regression coefficients and p-values).

My findings also produced several analytical results that suggest additional avenues of research for the future. In addition to political group identification, my study also collected preliminary observations about political group emotions, in line with intergroup emotions studies, which I detail in **Appendix C**, but a full study would require a different design focus and extended analysis. Emotions can be measured in a host of different ways and may necessitate a different methodology to assess how they factor in social media news use. Investigating the motivating role of intergroup emotions in political and social media contexts may prove particularly illuminating with regard to translating group identity into particular social media behaviours relative to political news encounters. Additionally, further research into sex/gender differences in both political group identification and in social media sharing and flagging behaviours would provide vital insight into the unexpected moderating effect of sex/gender my analyses revealed among my participants and which I explore further in **Appendix C**. That could involve discerning the existence of, and identification with, sex/gender-specific ‘subgroups’ within broader political groups (and potentially, discerning specific doctrinal differences between such subgroups as Democrat women and Democrat men concerning social media behaviour). Further research designs may need to consider subgroup relevance or salience in this kind of decision making.

The conceptual and empirical support provided by my findings most clearly highlights political group identification’s relationship with political ingroup news flagging and, to a somewhat less certain extent, with political outgroup news sharing. A social media user’s political group identification cannot be gleaned from the text of a news article or the purported trustworthiness of a news source, but among my study participants, it was increasingly observed as they expressed their willingness to flag genuine political news as fake. When users share content, flag a story as false, or click the various hearts, smiley faces, and other emoji that signify how they feel about a post, the social media platforms take in data. That data feeds into their algorithms and changes how stories are displayed, promoted, and moderated.⁷⁹⁷ But when users encounter political news on social media, they do so in relation to their identification with one or another political group.⁷⁹⁸ Across the groups I studied, more highly identified users tended to express greater willingness to flag genuine political news about their own groups as fake, regardless of whether they leaned Democrat or Republican. Flagging genuine stories as false in turn sends signals about news content to

⁷⁹⁷ Part II, ‘Platform Policies and Technical Capabilities’

⁷⁹⁸ Part III, ‘The Social Psychology of Political News and Social Media’

the platforms that distort what automated moderation systems ‘know’ about that content. In the next section, I discuss these implications further. As content moderation takes on legal consequence, such as through regulation and other means considered in Part V, these implications become increasingly serious with regard to free expression.

Implications

My experiment set out to test whether participants with higher political group identification would be more likely to 1) share negative fake news about the other group, and 2) flag genuine negative news about their own group as fake. The first hypothesis was partially supported by my findings that showed a significant association between group identification and sharing fake stories about the other group for one of my participant groups. The second hypothesis was fully supported by my findings showing a significant association in both groups between participants’ group identification and flagging genuine stories denigrating their own group as fake. In this section, I discuss those findings in light of the automated content moderation systems described in Part Two. My study helps show that when these systems rely on the ‘wisdom of the crowds’ to make determinations about news content, sharing and flagging data can reflect more than a desire for accuracy.

At the enormous scale of the largest platforms like YouTube, content policies are implemented through moderation processes that rely heavily on automation. These automated systems rely on data about the content to produce and act upon decisions about whether to promote or restrict a particular item. The vast user bases of these platforms provide essential data for those processes through their routine interactions with news content—liking and sharing stories, flagging or reporting questionable content, and other user data inputs that are fed into the platforms’ moderation algorithms. Accordingly, numerous studies have been undertaken to assess the efficacy of automated moderation systems in turning content data into appropriate responses to fake news.⁷⁹⁹ These studies have shown that automation exhibits a number of key limitations in accurately and consistently discerning genuine content from fake, and false positives—genuine content being treated by the automated systems as fake—are a key concern for both the platforms and advocates for free speech and the press.

The findings of my experiment suggest that the drives of social identity may colour the data about content relied upon in the status quo arrangement between social media companies, governments, and users. Social media companies operating at mass scale rely on

⁷⁹⁹ See Part II, ‘Platform Policies and Technical Capabilities’ – Evaluating Content Moderation at Scale

common technical approaches⁸⁰⁰ to automate the process of content moderation—including fake news content. Under these approaches, user flagging generates essential information about user-generated or user-shared content.⁸⁰¹ As described in Part Two, large social media companies use automated systems to simulate human moderation responses across huge volumes of content: the programmed rules of these systems produce certain outcomes, like promotion or demotion for various content, based on the data the system has about the content.⁸⁰² Computer scientists and legal scholars alike have pointed out that while automated decision-making systems can be designed for accountability, the decision rules and models used therein can nevertheless produce perverse results,⁸⁰³ particularly when they run into the limits of what these systems are actually capable of accomplishing.

Parts Three and Four of my Thesis have detailed one of these limits which is especially relevant in the context of news and political elections—the relationship of group psychology to individual decision-making. Elections create the perfect storm of heightened awareness of political group-membership, desire for reliable news information, and materially consequential outcomes. The experiment I conducted showed a significant relationship between my participants’ political identification and their willingness to flag genuine political news content as fake. My findings suggest that where social media companies rely on user flagging to identify certain content as fake news, a non-trivial portion of the flags may accordingly reflect motivated decision-making rather than the actual veracity of the stories. This source of error could be introduced in either of two key stages of the content moderation process: first, in identifying content to be subjected to moderation; and second, in training the models with which subject content is compared and acted upon.

In the first case, motivated flagging might not necessarily result in the direct curtailment of genuine political news and commentary. However, this kind of flagging in the first instance does put otherwise genuine expression into the moderation pipeline. Where additional moderation and action by platforms depends on feedback thresholds—reaching a particular critical mass of reports by users about a piece of content—motivated flagging effectively skews those thresholds lower for political news content than other news stories. This phenomenon compounds existing vulnerabilities to excessive moderation arising from

⁸⁰⁰ United Nations General Assembly ‘Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression’ (6 April 2018) UN Doc A/HRC/38/35

⁸⁰¹ Kate Crawford and Tarleton Gillespie, ‘What is a flag for? Social media reporting tools and the vocabulary of complaint’ (2016) 18(3) *new media & society* 410, 412-13

⁸⁰² Part II, ‘Platform Policies and Technical Capabilities’

⁸⁰³ Joshua A Kroll and others, ‘Accountable Algorithms’ (2017) 165 *U Pennsylvania L Rev* 633, 637-38 & 679-80

factors like the visibility of high-profile users (such as politicians or political activists).⁸⁰⁴ This also becomes particularly problematic where the user base of a social media platform may contain a dominant group that can drown out the input of smaller groups. Whether this increased risk of moderation translates into increased risk of suppression depends on whether the content models used by the social media companies then pass the subjected content on as genuine or not.. If those models can reliably discern between authentic and fake news stories, then the motivation for the content flag shouldn't matter: a flagged authentic story will pass muster and not receive inappropriate treatment by the platform. But if they can't, these automated systems may simply exacerbate existing tensions concerning political news, particularly to the detriment of disadvantaged groups.

In the second case, motivated flagging introduces the possibility of systemic suppression of genuine expression. Where motivated flagging is incorporated into the content models, false positives become part of the metric against which additional news stories are compared. As a result, automated content moderation systems will act on political news stories based in some part on characteristics unrelated to the genuineness of the content itself. Rather, the 'fakeness' signal is tied in part to the political group identity of the flagger. The potential for this source of error to enter content moderation models is substantial. In Part Two, I discussed how time-sensitivity surrounding elections news, among other factors, pressures social media companies to seek greater responsiveness from their algorithmic processes—even in the absence of specific regulatory requirements.⁸⁰⁵ Data generated by user flagging is the most abundant and cheaply available resource for the social media companies to mine in order to build their models. Unlike periodic training data from 'expert' annotators or third-party fact-checkers,⁸⁰⁶ user-generated data is constantly fed back to the platforms in real-time.

In the short, intense news cycles of elections, the continual feedback of user flagging provides social media companies with lots of especially timely signals about the content shared on their platforms. But these signals come with a substantial amount of noise which may not be discernible by computer models: what looks to the models like a binary decision of 'real or fake' in fact reflects a host of human judgments which are not impartial. Motivated flagging in the first instance can thereby expand the amount of genuine content incorporated into fake news models and likewise increase the number of false positives when the models are subsequently applied to additional news content in the second. Automated moderation

⁸⁰⁴ Discussed in Part II, 'Platform Policies and Technical Capabilities' – Evaluating Content Moderation at Scale

⁸⁰⁵ Part II, 'Platform Policies and Technical Capabilities' – Evaluating Content Moderation at Scale

⁸⁰⁶ While not the subject of my experiment, such expert training data itself likely also exhibits errors resulting from motivated decision-making for the same reasons discussed in Part III.

systems may thus worsen some of the problems discussed in Part One: timely access to genuine elections news may be hampered, and the availability of news content may take on a particularly political character—worsening polarization and frustrating public discourse.

The potential entanglement of social psychological phenomena in automated decision making processes, as described in this section, may in turn be worsened by regulation that requires large social media platforms to intensify their application of such systems. Similarly, transparency efforts may be frustrated by the inability of the platforms to discern or articulate motivated flagging in the data upon which they rely. In the next Part of my Thesis, I further discuss implications for free expression on social media under mass scale content moderation by examining its role in leading policy proposals.

Key Takeaways

- Social psychology research methods provided a platform-agnostic means for me to test my hypotheses that higher political group identification would correspond with greater willingness to share fake news stories about a political ‘outgroup’ and to flag (as fake) genuine news stories about a political ‘ingroup’
- My findings show that political group identity is associated with the decisions that large social media platforms ask users to make about the news content they encounter, particularly regarding news that reflects poorly on the user’s own group
- Automated ‘fake news’ moderation systems on large social media platforms may pose particular risks to the access of *genuine* political news by relying on criteria like user flags that represent more than the actual veracity of the stories involved

Part Five: Analysis and Conclusions about Proposals to Regulate Social Media

Social media sits amidst a news media landscape that stretches forth across centuries of cultural and technological development. As discussed in Part One, news has expanded from its earliest uses through technologies like the printing press, which facilitated the accurate and efficient reproduction of texts, and the telegraph, which dramatically reduced the cost and time required to distribute information. Democratic societies have come to feature news use as an important part of the political process, and social media has become an important venue for political news and discussion. The civic role of social media has been highlighted in numerous examples in recent years. Consequential political news has always been a matter of concern for politicians especially and society generally, and the enormous reach and influence of platforms like Facebook has attracted public scrutiny in light of the potential role of political news on the platforms in recent U.S. and U.K. elections. The phenomenon of ‘fake news’ in particular has produced no small amount of comment but relatively little agreement about how to define fake news or characterize the potential problems it poses for elections.

In the second Part of my Thesis, I undertook a deeper look at social media as a media system to help refine my research. Considering social media as a media system like any other, I traced a history of communications technology and media policy development. That history revealed key similarities and differences in how social media relates to broader socio-cultural spheres as compared with newspapers (for instance) including similarities (and extensions) of structural phenomena like industrial and infrastructural concentration and differences in terms of organizational culture and interactivity with legal systems. These key elements underpin my framework for evaluating whether novel policy proposals will appropriately engage with the social media system to achieve a particular civic function for communications on the platforms. I then took a closer look at several leading social media platforms to further characterize their use of content policies and automated moderation.

In my review of how automated content moderation is accomplished, I determined that a particular influence on content data likely remains opaque even to well-designed algorithms: the role of social psychology in users’ decision making about political news. Psychological research has elsewhere provided valuable insights for legal scholars and practitioners, and social psychology provides crucial understanding of how members of social groups make decisions about information. In Part Three, I discussed some of the principles of social identity theory, a framework that integrates how we see ourselves and others as members of distinct groups with the thoughts, feelings, and behaviours we exhibit in social contexts. The processes of social identification facilitate our social interactions and

drive us to develop and maintain positive and distinctive group identities. Past studies of media use have highlighted the role of group membership in our media ‘diets’—what information we seek out and what we find palatable. Social media provides a smorgasbord of information that can bear on our group identities and is no less prone to its users’ identity-driven motivations than past communications technologies like television. Moreover, the architecture of social media both accommodates group-level interactions (through sharing and flagging content) and relies on those interactions to provide the data necessary for content moderation. I surmised that these group-mediated interactions likely introduce some measure of error into the technical processes the platforms use to address whether particular user content is ‘fake’ news.

In order to test my assumption, I designed an experiment to assess the relationship of political group identification to decisions about social media news content, as described in Part Four. I measured my study participants’ level of identification with two American political parties and compared those levels with my participants’ reactions to genuine and fictitious political news stories they might see on social media platforms. I found substantial associations with group identification in the participants’ decisions to flag genuine ‘ingroup’ stories as fake. Those findings support my premise that social media user inputs about content may reflect social identity-driven motivations rather than actual discernment between genuine and fake political news. The implication for government and social media policy making is that continued reliance on automated content moderation will likely result in an on-going risk of false positives and frustration of efforts to make moderation more transparent: The ways we make decisions as social actors may be replicated but not necessarily revealed by the platforms’ algorithmic processes.

How then might the issue of consequential fake news content on massive social media platforms be effectively addressed? My empirical research provided me with key insights into unexpected risks to free expression carried by the status quo of platform algorithmic moderation and potential regulation that requires social media platforms to intensify the use of such systems. It does not, however, provide an answer to a question of such breadth—my experiment examined a technical structure specific to the media system of large platforms like Facebook. As such, my research plays a part in assessing how the use of automated content moderation may perform in answering that question. To fully consider what constitutes an effective policy position on social media fake news, I conduct the broader analysis and criticism contained in this Part. Here, I draw on media and law system analyses to assess policy choices and consider what trade-offs are involved in achieving a particular public function ‘vision’ for the social media platforms.

In this final Part of my Thesis, I critique a range of policy responses to fake political news under consideration in Britain and the United States. Public policy concerning social media tends to reflect specific perspectives on the role of government in communications systems, ranging from preferences for the platforms to manage user-generated content themselves to preferences for more direct state intervention to curtail unwanted political discourses. Across these perspectives, the social media platforms are typically envisioned as retaining a central role in the implementation of content policy, often owing to the scale and complexity of their operations and the technical expertise the platforms hold. In the first section of this Part, I explore how we conceptualize media policy as a means for promoting particular public functions of our media systems and describe specific perspectives relevant to social media as a local and global policy concern. In the section after that, I analyse potential 'allocation' decisions we might make in terms of assigning authority to our institutions relative to their particular characteristics. I conclude this Part with a discussion of what it means for us to make the various trade-offs inherent in addressing social media news content through law.

The first section of Part Five focuses on the American, European, and transnational perspectives which animate some of the key differences and similarities in policy approach across jurisdictions. This section also details significant public policy initiatives bearing on social media, including the Section 230 liability safe harbour and the proposed Online Safety Bill, as well as recent social media laws passed in Florida and Texas. Section 230 in particular looms large over discussions of social media policy: many of the major social media platforms are based in the U.S. and have developed their business models under its aegis. The Online Safety Bill and American state law developments represent the leading edge of public efforts to come to grips with how the platforms manage and distribute the content their users generate and how the platforms should answer to public scrutiny. The American system has produced growing tension between the powerful private platforms, lawmakers who push for platform transparency and accountability, and long-standing traditions of free expression that sharply demarcate appropriate government actions concerning media use. In Europe, a different array of cultural and social factors underpin lawmakers' determination to bring social media under the umbrella of public accountability while courting further growth and development of European social media businesses. Alongside these regional perspectives, national security services and multinational military alliances like NATO have called for policy changes to combat a wide array of threats arising from social media content. Each of these perspectives reflect specific interests and competencies that bear on where authority over social media content should be held, and likewise bear on how the different approaches they espouse may enhance or erode the kinds and quality of information the public can provide and receive over social media.

In the second section of this Part, I integrate my previous analysis and experimental results to critique prominent approaches to addressing fake news on social media. In doing so, I consider the extent to which policy decisions about responsibility for online fake news are decisions about how to allocate social power. The status quo and various proposals to regulate social media all represent choices about where to vest authority over content. Determining who is accountable for what content requires the empowerment of some agent to make decisions about social media content and hold another party to account. Likewise, particular interests are prioritized by the allocation of that power, be they public or private interests. Lastly, these allocations impact the exercise of our rights to free expression and a free press to greater or lesser extents based on what institutions are given such authority. As noted above, my research approach has uncovered particular implications where pursuing the public function goal of addressing fake news calls upon social media platforms to continue or extend their use of automated content moderation, as in the status quo and in regulatory or co-regulatory regimes. I further draw on my analysis of historical media system dysfunction and public response to go beyond my empirical research in considering alternative approaches for promoting the civic function of social media.

In the final section of this Part, I consider what it means for our society to make these various allocations of social power. Achieving a positive civic role for social media communications requires us to take stock of the values and principles bound up in the investment of such power in particular institutions. This socio-cultural exercise demands rigorous public reflection on which values and principles we prize over others in determining what policy to pursue—as well as recognition that other societies and cultures may determine their preferences differently. The far reach of the Internet makes this task no less simple, but the involvement of social media in our electoral processes makes this reflection and pursuant determination urgent in democratic societies.

Public Policy and the Public Sphere

In the following pages of this section, I return to the concept of the public sphere as a representation of the larger socio-cultural whole to which media systems relate. The social media system's relation to the public sphere in turn sheds light on the kinds of political and legal decisions might best help us shape its functioning—particularly when informed with greater insight into social media system factors like automated content moderation.

In Part Two, I outlined a media systems-oriented history of communications technologies and their legal treatment spanning the development of the printing press, newspapers, telegraphy, and the Internet. Printing was a truly revolutionary development

that enabled a new level of communication and intellectual culture. The printing press allowed for the mechanical reproduction of text, and thus many more people could read and talk about the exact same information. Rulers were naturally concerned about the content of that information and its bearing on their rule. Newspapers gained social influence as they became instrumental to readers, particularly those who wanted to decide for themselves what to do as a society about the conditions and events of their lives. Support for communications systems like the postal services helped them do that, and technologies like the telegraph promised to help them do that even more and faster.

The ideas of the Reformation and humanist revolutions of the 16th Century took on a “public life” through printing and publication in a way that was recognized at the time and has persisted as a feature of “the press” since then: the development of a “public sphere or domain in which opinions and ideas [are] aired, formulated, transmitted, and contested.”⁸⁰⁷ Yet historically, the “idea that men have to go forth and study the world in order to govern it” developed only very recently compared to other aspects of democratic theory and political science, essentially emerging in relation to the communications technologies and social conditions which enabled it.⁸⁰⁸ The accountability concerns surrounding social media news use are intimately bound up in contemporary cultural conceptions of news media itself. The modern notion of the newspaper “carries with it the connotation, often nurtured by the press itself, of being an active and significant participant in society,”⁸⁰⁹ a role which invites a more particular level of accountability than other narrative forms.

Social media is no different in the sense of providing us with a shared body of knowledge about distant events. Nor does social media mark a break from the previous social order achieved through liberalism, instead embodying its latest technological expression. By contrast, the printing press supported popular desires to reliably share information outside the political and religious structures of the late-Medieval period that had previously monopolized the production of accounts of what the world around us was like. In this sense, complaints about too much social media user participation in the public sphere and too many diverse accounts about news events are complaints about liberalism itself,

⁸⁰⁷ Denis McQuail, *Media Accountability and Freedom of Publication* (OUP 2007) 32

⁸⁰⁸ Walter Lippmann, *Public Opinion* (Simon & Schuster 1997) 201

⁸⁰⁹ McQuail (n 807) 34-35; cf. Gaetano Mosca, *The Ruling Class* (McGraw-Hill 1939) ch 5, §§ 6 & 7

(describing how the balance of social forces is not merely formal or legalistic: “if one political institution is to be an effective curb on the activity of another it must represent a political force—it must, that is, be the organized expression of a social influence and social authority that has some standing in the community, as against the forces that are expressed in the political institution that is to be controlled”)

echoing mid-20th Century concerns about an excess or crisis of democracy brought on by too much public and press involvement in politics.⁸¹⁰

Today's Internet distribution of news and opinion does mark the zenith of immediacy and availability, particularly thanks to cheaply available Internet-enabled mobile phones. As discussed previously, both official censorship by prior restraint of publication and the official policing of media distribution would be substantially reduced in scope by new forms of government that derived their authority (at least in part) through democratic participation and public discourse about worthy government undertakings. But as the locus of social, economic, and cultural power shifted away from the aristocratic state into private hands—the eventual industrialized press and telegraph monopolies—the consequences of concentrated power over news distribution remained as a threat to popular democracy. The public state's power would in turn be marshalled to protect access to and participation in the emerging public sphere. As it had in the past, the concentration of private power over news distribution has now become a vital concern regarding social media. These histories provide key insights into how the broader function of social media may change if certain public policy changes are enacted to re-allocate authority from the platforms to various other institutions.

Many of the largest social media platforms do hold themselves out as the new public square or the place where society comes together in conversation, as detailed in Parts One and Two.⁸¹¹ Some platforms like YouTube and Facebook provide a very broad range of content and features, serving not only as town square but also shopping mall, movie theatre, arcade, etc.—while Twitter positions itself as more focused on serious news and political discussion—and still others like Instagram and Pinterest take on particular interest niches in the way magazines do. As a technology in their own right, the social media platforms add relatively little new to the social or individual production and consumption of discourses⁸¹²—the most direct modern analogue to the printing press, in terms of elevating the recording, reproducing, and experiencing of information, might be essential hardware like the personal computer or Internet-enabled mobile phone. In many ways, the industrial-scale social media platforms may tend to degrade the production and consumption of news narratives, whether through limited and prescriptive formats, or architectures which promote endless scrolling and dopamine chasing rather than establishing truths. But as a distribution channel, social media can nevertheless tend to approximate the ideal of the

⁸¹⁰ Discussed *infra*, – American Perspectives, – European Perspectives

⁸¹¹ Part I, 'Social Media News Use' and Part II, 'Platform Policies and Technical Capabilities' – Facebook, – Twitter

⁸¹² As observed in a tweet by actor/recording artist and social media enthusiast Ice T, "Twitter is a Chat Room... Nothing more."—a pithy summation of the 'micro-blogging' platform that repackaged weblog posting for mobile phone-based production and consumption; Ice T, 'Status' (23 February 2019) <<https://x.com/FINALLEVEL/status/1099389177483878401>> accessed 30 November 2023

public sphere in some important ways by leveraging the incredible reach and interactive possibilities of the Internet. Yet the interactive aspect of social media platforms itself remains an intensely mediated aspect. Unlike people having a chat over the newspaper and a coffee, the experience is curated not just in the received information but also in the subsequent conversation where it is still subject to platform moderation.

Social media companies are in this sense not mere ‘gatekeepers’ like Internet service providers or search engines. As content aggregators, the social media companies hold regulatory force over the information on their distribution platforms, and they do so in ways which have not been permitted to private carriers since the heights of control achieved by Western Union. Nor has such a level of control even been permitted to public regulators, which in the U.S. have been bound by the First Amendment and in the U.K. by Article 10 of the Human Rights Act.⁸¹³ Section 230, and to a relatively lesser extent the EC Directive hosting defence, set the stage for the social media industry to develop its own means for regulating the transmission of news and other information by its users—including the mass-scale automated content moderation systems in use today. The underlying anxiety in questions of what is to be done about social media news content arises from the mounting recognition that this power now resides in platform algorithms, and from uncertainty over who should be trusted to hold that power if it can be taken back.

The role of media in the public sphere also helps reveal how media industry concentration and political entanglement contribute to consequential distortions in public discourse. The historical concept of a public sphere in which communication is freely mediated by independent producers in many ways seems to stand at odds with the myriad pecuniary and political interests of today’s conglomerated newspaper publishers and 24-hour cable news giants, the competing interests of the subjects of news stories (especially politicians), and of readers and viewers that consume such narratives for diverse and idiosyncratic reasons. The continued political attention towards the media (including social media), on the other hand, reflects how “the immediacy and actuality of the news and opinion, its rapid diffusion at times of unrest, its orientation to, and availability to, a much larger public”⁸¹⁴ has intensified the danger of certain popular narratives to society—or particularly to the authorities themselves.

The First World War marks a key turning point in the relationship between government and the press that continues to shape how political interests view media.⁸¹⁵ During the late nineteenth century, newspapers mirrored societal shifts toward greater

⁸¹³ Discussed further below

⁸¹⁴ McQuail (n 807) 35

⁸¹⁵ See also **Appendix D** for discussion of Espionage Act cases from the U.S.

industrialization and, accordingly, consolidation and concentration.⁸¹⁶ At that time, the press was not necessarily aligned with government interests—at least not so far as it was rooted in the popular and radical press traditions of the 19th Century that decried the aristocracy (whether the actual kind in Europe or the so-called “robber barons” in America). That began to change in the U.S. by the late 1890s, particularly with William Randolph Hearst, the newspaper magnate who shared the racialized, imperialist ambitions of policy planners in the McKinley and Roosevelt administrations.⁸¹⁷ Hearst’s dominant position in the newspaper industry allowed him to push pro-war narratives on a mass scale.⁸¹⁸ When an American ship, the *USS Maine*, exploded and sank in the Havana harbour of Spanish controlled Cuba, Hearst “proclaimed in his *New York Journal* that the sinking of the *Maine* was ‘not an accident,’ an idea fed to him by Theodore Roosevelt, [and] ‘Remember the *Maine*’ fever enveloped the nation.”⁸¹⁹ Thomas Edison’s motion picture company and others, “operating with state sanction,” developed and exhibited “thrilling” movies spectacularly depicting the *Maine* disaster.⁸²⁰ The Spanish-American War ensued, and the utility of favourable news coverage was thoroughly impressed upon American politicians.

The American public likewise had little appetite for U.S. involvement in World War I, and Woodrow Wilson was re-elected President in 1916 with the campaign slogan, “He Kept Us Out of War.”⁸²¹ Wilson’s almost immediate about-face on the war *after* winning re-election necessitated closer coordination with the press than in the Hearst venture and a more sophisticated approach to change public opinion about a seemingly distant conflict.⁸²² Wilson created the Committee on Public Information, headed by journalist George Creel: a massive multimedia blitz launched to shape public conversation about the war, including trained speech deliverers, pamphlets, cartoons, magazines, and movies depicting the Central Powers as inhuman monsters.⁸²³ And despite the generally pacifist, isolationist, and pro-labour attitudes of Americans beforehand, Wilson successfully asked the Congress to declare war in April 1917.⁸²⁴ The President, the press, and the war effort proceeded in lockstep. Subsequent technological advances in industrialised communications did not substantially change the media system factors that facilitated such top-down influences on public

⁸¹⁶ Eric Louw, *The Media and Cultural Production* (SAGE 2001) 49-50; as discussed in Part Two, structural changes to the media system (i.e., industrial concentration) likewise interacted with changes in the law system (adoption of regulatory bureaucracies)

⁸¹⁷ Christopher Sharrett, ‘9/11, the Useful Incident, and the Legacy of the Creel Committee’ (2004) 43(4) *Cinema Journal* 125, 126

⁸¹⁸ *ibid*

⁸¹⁹ *ibid* 126-127

⁸²⁰ *ibid* 127

⁸²¹ Woodrow Wilson House, ‘Election: 1916’ (November 19, 2020)

<<https://www.woodrowwilsonhouse.org/election-1916/>> accessed 6 March 2022

⁸²² Sharrett (n 817) 127-128

⁸²³ *ibid*

⁸²⁴ *ibid* 128

discourse. As observed by Louw, the television would elevate the “corporatized, industrialized mode of communication” in the U.S. and U.K. alike, rounding out the mass media communications institutions able to “set the agenda for what people discussed.”⁸²⁵

Social media companies exhibit similar characteristics of industrial concentration in their centralised control over information distribution, and of government involvement in their close relationships with security services like the FBI and DHS in the U.S. and NATO in Europe. As with media industries in the past century, these characteristics present state actors with the opportunity to exert enormous influence over public conversation about foreign policy and war via social media. Distant events involving people and places we don't know still represent the furthest limits of our testable knowledge.⁸²⁶ Lippmann observed that these “limitations on our access to that [distant] environment combine with the obscurity and complexity of the facts themselves to thwart clearness and justice of perception, to substitute misleading fictions for workable ideas, and to deprive us of adequate checks upon those who consciously strive to mislead.”⁸²⁷ The promise of access to social media should tend to militate against that by allowing for more people to provide more information about news events and thus give social media users the chance to rely less on generalizations and stereotypes to fill in the gaps in their understanding of the world around them. But where government agencies are able (or enabled by the platforms) to restrict access to social media news sources for reasons of ‘national security’, social media users are pushed back into the lower-information state that facilitates top-down shaping of the discourse around such events.

In recent years, many governments around the world have taken extraordinary measures to manage their citizens' use of the Internet in advance of elections, including steps taken to restrict or block access to various social media platforms and messaging services, despite condemnation by the U.N. Human Rights Council of such measures as excessively broad.⁸²⁸ In their wide-ranging survey of fake news policies, Helm and Nasu categorized regulatory approaches based on their intrusiveness relative to international law norms about free expression, from ‘information correction’ to ‘content removal or blocking’ to ‘criminal sanction’ and liability regimes.⁸²⁹ Laws restricting “fake news,” “online falsehoods,” “false information,” and “disinformation” during election campaigns have been introduced in

⁸²⁵ Louw (n 816) 38

⁸²⁶ Lippmann (n 808) 95

⁸²⁷ *ibid* 49

⁸²⁸ United Nations General Assembly ‘Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression’ (6 September 2016) UN Doc A/71/373, 9-10

⁸²⁹ Rebecca K Helm and Hitoshi Nasu, ‘Regulatory Responses to “Fake News” and Freedom of Expression: Normative and Empirical Evaluation’ (2021) 21 Human Rights Law Review 302

parliaments from Singapore to France.⁸³⁰ Malaysia's Anti-Fake News Act 2018, for example, sought to criminalize the malicious creation, publication, or dissemination of fake news, including stiff penalties and imprisonment for offenders.⁸³¹ Still other countries such as China and Saudi Arabia have pursued restrictive policies in response to burgeoning Internet communications, in particular relying on filtering and blocking technologies to restrict access to certain online content.⁸³²

Under German law, Internet platforms with more than 2 million registered users in Germany must remove certain illegal content within 24 hours of notice, including, under German criminal law ("*NetzDG*"), Holocaust denial⁸³³ and Nazi propaganda,⁸³⁴ and must regularly report on their procedures.⁸³⁵ The U.K. Government has put forward a proposal for a statutory duty of care for tech companies, with oversight by an independent regulator who will promulgate "codes of practice" relating to online safety, aiming to provide clarity for companies "especially in the context of harmful content or activity that may not cross the criminal threshold but can be particularly damaging to children or other vulnerable users."⁸³⁶ And in the United States, Senator Warner, who co-led the Senate Intelligence Committee investigation into Russian social media influence on the 2016 U.S. presidential election, wrote about the need to put "common-sense rules of the road" in place for social media companies to start "taking responsibility for their platforms and the potential for their abuse."⁸³⁷

The details of the proposals included in this section highlight some of the aims being pursued by lawmakers regarding the civic function of social media. Differences between the proposals presented here also reveal the range of normative values held about how that function should be achieved and how other values like social cohesion or political liberty are being weighed relative to the purported harms of fake news. Furthermore, some proposals like the Online Safety Bill or SB 7072 aim to bring certain aspects of content moderation policy in line with government expectations. These proposals are 'co-regulatory' in the sense

⁸³⁰ Joseph M Fernandez, 'Malaysia's Anti-Fake News Act' (2019) 25 *Pacific Journalism Review* 173, 174-77

⁸³¹ *ibid* 179

⁸³² John Palfrey, 'Four Phases of Internet Regulation' (2010) in 77(3) *social research* 981, 985-96

⁸³³ § 130 StGB [Germany, 1998] (translation available at <https://www.gesetze-im-internet.de/englisch_stgb/englisch_stgb.html>) accessed 25 February 2020

⁸³⁴ § 86 StGB [Germany, 1998] (translation available at <https://www.gesetze-im-internet.de/englisch_stgb/englisch_stgb.html>) accessed 25 February 2020

⁸³⁵ Google, 'Removals under the Network Enforcement Law' (*Google.com*, 31 December 2019) <https://transparencyreport.google.com/netzdg/youtube?hl=en_GB> accessed 25 February 2020

⁸³⁶ Department for Digital, Culture, Media and Sport, *Online Harms* (White Paper, CP 57, April 2019) 7 (hereinafter '*Online Harms White Paper*')

⁸³⁷ Mark Warner, 'Warner: It's past time to learn from failures, adapt our laws to the internet age' (01 August 2018) <<https://www.warner.senate.gov/public/index.cfm/2018/8/warner-it-s-past-time-to-learn-from-failures-adapt-our-laws-to-the-internet-age>> accessed 28 July 2022

that they rely on the platforms to use various techniques (as discussed in Part Two) to meet public function goals like the free flow of news information from designated media outlets. Accordingly, the effectiveness of platform content moderation techniques bear directly on the potential effectiveness of these regulatory proposals. I revisit the implications of reliance on the platforms in the section after this.

American Perspectives

Warner and other Senators had previously introduced legislation proposing transparency rules for online political advertisements, which Facebook and Twitter subsequently used in developing their political advertising policies.⁸³⁸ Warner further proposed measures aimed specifically at countering foreign disinformation, protecting consumers, and promoting competition and innovation.⁸³⁹ Such measures would include the clear labelling of bots and “inauthentic” accounts, reforming tort liability for online defamation and invasion of privacy, improving political advertising disclosures, and promoting media literacy among the general public.⁸⁴⁰ Senator Elizabeth Warren of Massachusetts included fighting “digital disinformation” as part of her 2020 presidential election campaign, emphasizing that “disinformation campaigns” threatened American democracy and specifically targeted minorities.⁸⁴¹ Members of the U.S. House of Representatives have also pressed the CEOs of social media companies regarding accountability for “misinformation and disinformation plaguing online platforms,” declaring that “industry self-regulation has failed.”⁸⁴²

Scholarship supporting the debate on how to tackle such threats to American democracy has deep roots. In 1787, Madison argued that the aim of breaking and controlling factionalism, which he viewed as the source of “instability, injustice, and confusion” ultimately fatal to popular government, should be chief among the aims of the post-Revolutionary government.⁸⁴³ Madison identified two methods of addressing factionalism:

⁸³⁸ Warner (n 837)

⁸³⁹ Mark Warner, ‘Potential Policy Proposals for Regulation of Social Media and Technology Firms’ (White Paper, August 2018) <<https://www.scribd.com/document/385137394/MRW-Social-Media-Regulation-Proposals-Developed>> accessed 28 July 2022

⁸⁴⁰ *ibid*

⁸⁴¹ Elizabeth Warren, ‘Fighting digital disinformation’ (*Warren for President*, 2020) <<https://elizabethwarren.com/plans/fighting-digital-disinformation>> accessed 28 July 2022 (Warren extensively cites the New Knowledge report discussed in the previous section)

⁸⁴² US House Committee on Energy & Commerce, ‘E&C Committee Announces Hearing with Tech CEOs on the Misinformation and Disinformation Plaguing Online Platforms’ (18 February 2021) <<https://energycommerce.house.gov/newsroom/press-releases/ec-committee-announces-hearing-with-tech-ceos-on-the-misinformation-and>> accessed 28 August 2022

⁸⁴³ James Madison, ‘The Same Subject Continued: The Union as a Safeguard Against Domestic Faction and Insurrection’ *The Federalist Papers No. 10* (23 November 1787)

first by removing its causes, and second by controlling its effects.⁸⁴⁴ In dealing with the causes, Madison warned against curtailments of liberty on the grounds that liberty is “essential to political life” despite the nourishment it gives to competing opinions and interests.⁸⁴⁵ Madison also suggested that the other means of removing the causes of faction, “by giving to every citizen the same opinions, the same passions, and the same interests,” would be “impracticable” in the face of human fallibility and diversity.⁸⁴⁶ Accordingly, Madison addressed his analysis to ameliorating the effects of societal division, prescribing the republican form of government which the United States adopted in its Constitution.⁸⁴⁷ Yet Madison’s diagnosis of the “factious spirit” tainting America’s public administration⁸⁴⁸ would be revived in later analyses, particularly in the Twentieth Century amidst the development of mass media and mass communications technologies.⁸⁴⁹

Huq and Ginsberg, for example, in their sweeping 2018 survey of threats to constitutional democracy, argue that the broad protections of the First Amendment’s free speech clause may contribute (among a plethora of other factors) to “democratic backsliding” by shielding politicians who criticize their opponents or the media and precluding the adoption of broad laws punishing false speech on social media as in other countries.⁸⁵⁰ Such views echo the conclusions of earlier analyses, like Huntington’s attribution of the loss of trust and confidence in governmental leaders and institutions to the “democratic surge” of “a more politically active citizenry, which developed increased ideological consistency on public issues” like race and social welfare in the late 1950s and early 1960s.⁸⁵¹ Huntington noted as significant that “only the leadership of the press and television news enjoyed more confidence in 1973 than they had in 1966” as compared to declines among government and other major social institutions.⁸⁵² Moreover, Huntington described the national television networks, news magazines, and mass-circulation newspapers as the “most notable new source of national power in 1970, as compared to 1950,” in his analysis of the diffusion of control away from the more concentrated centres of power supporting U.S. authority post World War II.⁸⁵³ Huntington points to this change in political relationship as a potential explanation for the increase in citizens’ perceptions of national politics as central to their

⁸⁴⁴ Madison (n 843)

⁸⁴⁵ *ibid*

⁸⁴⁶ *ibid*

⁸⁴⁷ *ibid*

⁸⁴⁸ *ibid*

⁸⁴⁹ As discussed in Part II, ‘Social Media and Media Systems’

⁸⁵⁰ Aziz Huq and Tom Ginsberg, ‘How to Lose a Constitutional Democracy’ (2018) 65 UCLA Law Review 78, 153-57

⁸⁵¹ Samuel P Huntington, ‘The United States’ in *The Crisis Of Democracy* by Michel Crozier, Samuel P Huntington, & Joji Watanuki (1975 NYU Press) 59, 76-78

⁸⁵² *ibid* 80

⁸⁵³ *ibid* 97-100

own personal lives.⁸⁵⁴ In today's intensively connected social media ecosystem, advocates like Wu have called for a return to the media environment of the 1950s through appropriate policies for online platforms.⁸⁵⁵

The outcomes of *Reno* and several cases discussed earlier in this Thesis⁸⁵⁶ remain intensely relevant today because they outline what kinds of legislative approaches to social media regulation may be constitutionally viable in the U.S. However, the decisions only tell part of the broader story about how control over the social influences of social media is actually balanced between public and private hands. Section 230, the key survivor of *Reno*'s repudiation of legislative control over Internet content, has its own important history that reveals much about where regulatory authority would rest over the decades following *Reno*. Section 230 introduced a massive warp into the fabric of American communications law by pulling Internet publisher-distributors out of the established legal process. Where newspapers faced liability for what they published, "interactive computer service" providers would not,⁸⁵⁷ and where telephone companies could not restrict constitutionally protected communication, these Internet intermediaries would be shielded from liability for access discrimination or censorship.⁸⁵⁸

The international character of the Internet has necessitated other nations' development of their own regimes to address its myriad uses,⁸⁵⁹ and in Europe, many nations have taken an approach more similar to the U.S.'s Digital Millennium Copyright Act⁸⁶⁰ than Section 230.⁸⁶¹ Such provisions grant protection from liability on the condition that the service provider shall maintain a certain removal from involvement with infringing information transmissions and take action to remove such content upon notice thereof.⁸⁶² In line with the June 2000 European Directive on electronic commerce,⁸⁶³ the U.K. enacted the Electronic Commerce (EC Directive) Regulations 2002,⁸⁶⁴ which include conditional liability provisions for "information society service" providers acting as mere conduits for a

⁸⁵⁴ Huntington (n 851) 110-13

⁸⁵⁵ Tim Wu, 'Is the First Amendment Obsolete?' in *The Free Speech Century* (eds Geoffrey R Stone & Lee C Bollinger, Oxford UP 2019) 272, 290-91; Wu's views are not without critics—see, e.g., Matt Taibbi, 'A Biden Appointee's Troubling Views On The First Amendment' (*TK News*, 23 March 2021) <<https://taibbi.substack.com/p/a-biden-appointees-troubling-views>> accessed 28 July 2022

⁸⁵⁶ See Part II, 'Social Media and Media Systems'

⁸⁵⁷ § 230(c)(1)

⁸⁵⁸ § 230(c)(2)

⁸⁵⁹ McQuail (n 807) 111

⁸⁶⁰ Digital Millennium Copyright Act of 1998, 105 PUB. L. 304

⁸⁶¹ Tarleton Gillespie, 'Governance of and by platforms' in Jean Burgess, Alice Marwick, and Thomas Poell (eds), *The SAGE Handbook of Social Media* (SAGE 2018) 6

⁸⁶² *ibid* 6-7

⁸⁶³ Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce)

⁸⁶⁴ Electronic Commerce (EC Directive) Regulations 2002

transmission,⁸⁶⁵ providers merely caching transmissions,⁸⁶⁶ and providers who host or store information⁸⁶⁷ generated by their users. The EC Directive also places the burden on prosecutors to overcome these defences beyond a reasonable doubt once sufficiently raised by a service provider.⁸⁶⁸ The hosting defence⁸⁶⁹ is particularly relevant for providers like YouTube and other social networks that act as “content aggregators.”⁸⁷⁰ The U.K. approach accommodates a broad range of issues arising from user-generated content, including copyright, defamation, privacy, and incitement, though each issue draws in further enforcement considerations in each of those areas of law, though generally may be mitigated by providers through adequate ‘notice and take-down’ procedures.⁸⁷¹ Gillespie notes that while such procedures can generate challenges for Internet intermediaries (and may be prone to abuse—the ‘heckler’s veto’ by discomfited fellow users) these conditions are, like the broader Section 230 protections, nonetheless “profoundly advantageous” and “far preferable for platforms than being held liable for what their users post.”⁸⁷²

U.S. lawmakers from both the Democrat and Republican parties have “signalled their support to modify Section 230” in pursuit of various public policy goals, including the reduction of online hate speech, incitement, and elections interference.⁸⁷³ ‘Fairness’ has also been centred as a key aspect of proposed legislation like the Ending Support for Internet Censorship Act, which would make the liability protection of Section 230 conditional upon a social media company’s certification by the Federal Trade Commission that the platform does not moderate user generated content in a manner that is “biased against a political party, candidate, or viewpoint.”⁸⁷⁴

As described thus far, the American perspective on social media regulation may seem to be extremely disordered. In the early days of the republic, Madison made a staunch (if begrudging) defence of the benefit of liberty to public life—despite the resulting difficulties of societal division. In the 20th Century, as new broadcast communications media diffused power outwards, American governments curtailed that liberty—acting in the interest of the public while doing so. Concern that online communications could displace regulated media

⁸⁶⁵ EC Directive Regulations, Reg. 17

⁸⁶⁶ *ibid* Reg. 18

⁸⁶⁷ *ibid* Reg. 19

⁸⁶⁸ *ibid* Reg. 21

⁸⁶⁹ *ibid* Reg. 19

⁸⁷⁰ Steve Holmes and Paul Ganley, ‘User-generated content and the law’ (2007) 2(5) *J Intellectual Property Law & Practice* 338, 39-40

⁸⁷¹ *ibid* 340 & 342-44

⁸⁷² Gillespie (n 861) 7; discussed also in Part V, ‘Analysing Allocation Decisions’

⁸⁷³ Kari Paul, ‘Twitter and Facebook CEOs testify on alleged anti-conservative bias’ (*The Guardian*, 17 November 2020) <<https://www.theguardian.com/us-news/2020/nov/17/twitter-facebook-senate-hearing-anti-conservative-bias>> accessed 28 July 2022

⁸⁷⁴ Ending Support for Internet Censorship Act, S. 1914 (116th Congress 1st session, 19 June 2019) § 2(a)(1)

competed with anticipation of the Internet’s innovative potential. The compromise, Section 230, allowed internet intermediaries like the social media platforms to self-regulate without worrying about liability for the content their users generated. Fake news, as a policy concern, thus invites proposals for incentivizing certain moderating behaviour by the platforms or creative rationalizations for intervening directly. But from the Madisonian perspective that prioritizes liberty, the perspective in a sense baked into the U.S. Constitution, the answer is to protect the expression of all perspectives, while the power of the state should instead be directed towards ameliorating the effects of societal division through the promotion of transparency and fair trading between the social media platforms and the users.

As noted by the U.S. Department of Justice, complaints of bias against the social media platforms are difficult for third parties to evaluate due to the lack of comprehensive data about content policy enforcement decisions.⁸⁷⁵ Transparency in content moderating decisions has become a centrepiece in laws passed by two of the largest U.S. states, Florida and Texas. Florida’s Senate Bill 7072 (“SB 7072”)⁸⁷⁶ created a requirement for social media platforms of a certain size to “publish the standards, including detailed definitions, it uses or has used for determining how to censor, deplatform, and shadow ban,”⁸⁷⁷ as well as to apply such standards in “a consistent manner among its users.”⁸⁷⁸ The Florida law also requires such social media platforms to provide users prompt notice explaining enforcement actions, an annual statement about the use of certain categories of algorithms used in content moderation (and with regard to which users must be allowed to opt out), and a period of continued access or retrieval for a user’s data or content when a user is “deplatformed.”⁸⁷⁹ SB 7072 further specifies that the notice to users include, in addition to the explanation of the reason for any censorship, an explanation of how the platform became aware of the content, including the use of algorithms to identify objectionable content.⁸⁸⁰

The civil cause contemplated by the Florida legislation, “an unfair or deceptive act or practice,” reflects the law’s overall consumer protection-orientation.⁸⁸¹ A private right of action is available where a platform fails to apply content standards in a consistent manner among its users or where the platform takes certain enforcement actions without providing notice to the user, with possible remedies including statutory damages, injunctive relief, and

⁸⁷⁵ U.S. Department of Justice, ‘Section 230 – Nurturing Innovation or Fostering Unaccountability?’ (June 2020) 25 <<https://www.justice.gov/file/1286331/download>> accessed 28 July 2022

⁸⁷⁶ Senate Bill No. 7072, LAWS OF FLORIDA, Chapter 2021-32 (effective 01 July 2021) (“SB 7072”)

⁸⁷⁷ *ibid* § 501.2041 (2)(a); SB 7072 defines ‘censor,’ ‘deplatform,’ and ‘shadow ban’ as (respectively) actions to restrict or edit content, temporarily or permanently ban a user, and limit or eliminate exposure of a user or user’s content to others, § 501.2041 (1)

⁸⁷⁸ *ibid* § 501.2041 (2)(b)

⁸⁷⁹ *ibid* §§ 501.2041 (2)(d), (e), (i)

⁸⁸⁰ *ibid* §§ 501.2041 (3)(c), (d)

⁸⁸¹ *ibid* § 501.2041 (2)

costs and attorney fees in certain circumstances.⁸⁸² Civil or administrative actions concerning other suspected violations are reserved to the State’s consumer protection authority.⁸⁸³ The drafters of the legislation additionally included language in the Bill characterizing the rationale for the law. The preamble to SB 7072 makes specific reference to social media platforms as “the new public town square” and “as important for conveying public opinion as public utilities are for supporting modern society” and accordingly “should be treated similarly to common carriers.”⁸⁸⁴ This civic-minded rationale underpins the law’s additional protections with regard to moderation for qualified electoral candidates during campaigns and more generally for “journalistic enterprises.”⁸⁸⁵ The electoral candidate provision specifically prohibits social media platforms from prioritizing or de-prioritizing content by *or about* qualified candidates for state or local office during the period of their candidacy (excepting paid promotion).⁸⁸⁶ SB 7072 further prohibits deplatforming a qualified candidate during such user’s candidacy.⁸⁸⁷

The Texas law, House Bill 20 (“HB 20”),⁸⁸⁸ adopts a framework comparable to the Florida law but with some noteworthy differences. HB 20 includes similar language in its preamble, declaring that “social media platforms function as common carriers” and that they are “public forums for public debate.”⁸⁸⁹ But the introductory language in HB 20 goes further than the Florida text, including explicit reference to a “public interest” affecting social media platforms and also to their enjoyment of federal government support and market dominance.⁸⁹⁰ Furthermore, HB 20 includes as a premise the notion that individuals have a “fundamental interest in the free exchange of ideas and information, including the freedom of others to share and receive ideas and information.”⁸⁹¹ Such language hews more closely to European notions of a positive right to receive and impart information, rather than traditional American negative expressions of the right to free speech as against government interference.⁸⁹² On this premise, HB 20 prohibits social media platforms from censoring “a user, a user’s expression, or a user’s ability to receive the expression of another person” based on the viewpoint of the user or the viewpoint represented in the user’s expression (or

⁸⁸² SB 7072 § 501.2041 (6)

⁸⁸³ *ibid* §§ 501.2041 (5), (8)

⁸⁸⁴ *ibid* §§ 1(4)-(6)

⁸⁸⁵ *ibid* §§ 501.2041 (2)(h), (j); SB 7072 defines a ‘journalistic enterprise’ in terms that typically contemplate large entities, *i.e.*, FCC licensed broadcasters, cable television channels, and major circulation newspapers, § 501.2041 (1)(d)

⁸⁸⁶ *ibid* § 501.2041 (2)(h)

⁸⁸⁷ *ibid* § 106.072(2)

⁸⁸⁸ House Bill 20, TEXAS BUSINESS & COMMERCE CODE, Subtitle C, Title 5, Chapter 120, and TEXAS CIVIL PRACTICE AND REMEDIES CODE, Title 6, Chapter 143A (effective 09 September 2021) (“HB 20”)

⁸⁸⁹ *ibid* § 1(3)

⁸⁹⁰ *ibid* §§ 1(3), (4)

⁸⁹¹ *ibid* § 1(2)

⁸⁹² For background on the development of U.S. ‘public interest’ regulatory standards, see **Appendix D**

that of another person),⁸⁹³ excepting specific authorization for such censorship by federal law, cases of child sexual exploitation, incitement, threats due to protected status, or other unlawful expression, as well as content filtering choices requested by the user.⁸⁹⁴

Like SB 7072, the Texas law calls for public disclosures by social media platforms of a certain size. HB 20 requires disclosures about content targeting and promotion, moderation, algorithms and ranking procedures, and details on how users' performance data is used, emphasizing the importance of users' ability to "make an informed choice" about using the platform's services.⁸⁹⁵ HB 20 requires platforms to publish an "acceptable use policy" informing users of the platform's content policy⁸⁹⁶ and a transparency report every six months including the number of alerts about illegal or policy-violating content and the number of times the platform took action with regard to such content (contemplating a broader range of enforcement actions compared to SB 7072).⁸⁹⁷ The Texas law also sets guidelines for the reporting, requiring enforcement actions to be paired with the relevant rule violated and description of the source of the alert (including automatic detection tools like algorithms) to be provided.⁸⁹⁸ HB 20 further requires that social media platforms provide notice to users concurrent with any content removals, including an explanation of the reason why and an opportunity for appeal.⁸⁹⁹ Unlike the private right of action provided by the Florida statute, the Texas law may be enforced only by the State Attorney General, and the relief provided is injunctive.⁹⁰⁰ The Texas law provides no special provisions with regard to political elections.

Following the enactment of each of the Texas and Florida content moderation laws, two trade associations representing social media companies, NetChoice and the Computer and Communications Industry Association, sued the relevant state officials in Texas⁹⁰¹ and Florida⁹⁰² to enjoin the enforcement of the new laws. In the federal appeals court reviews of the cases, the court hearing the Texas case reached significantly different conclusions about Supreme Court precedent from the court hearing the Florida case.⁹⁰³ These differences turn on two fundamental questions that carry substantial implications for any future American efforts to address fake news on social media: where does internet communications

⁸⁹³ HB 20 § 143A.002

⁸⁹⁴ *ibid* §§ 143A.006(a), (b)

⁸⁹⁵ *ibid* §§ 120.051(a), (b)

⁸⁹⁶ *ibid* § 120.052

⁸⁹⁷ *ibid* § 120.053(a)

⁸⁹⁸ *ibid* § 120.053(b)

⁸⁹⁹ *ibid* § 120.103(a)

⁹⁰⁰ *ibid* § 120.151

⁹⁰¹ *NetChoice, LLC v. Paxton*, No. 21-51178 (5th Cir. 2022) ("*Paxton*")

⁹⁰² *NetChoice, LLC v. Att'y Gen. of Fla.*, No. 21-12355 (11th Cir. 2022) ("*Florida*")

⁹⁰³ *Paxton* (n 901) 80-89

technology fit in the historical landscape of media regulation, and which interests of the public are relevant in making public policy about social media? I discuss these cases in detail in Part Five of this Thesis in context with the relevant precedents of American free speech law. In light of the split between the appellate courts, the Supreme Court may soon look at whether the First Amendment prohibits a State from requiring social media companies to host users' communications or regulating how they do so, and whether the First Amendment prohibits a State from requiring social media companies to notify and provide an explanation to users about censorship decisions.⁹⁰⁴

Section 230 and the more recent state law initiatives reflect the tension that has arisen amidst American leering towards direct government intervention in communications and the power over communications held by the social media platforms. The existing framework under Section 230 and the new laws each envision a continued role for the platforms in managing the content users generate. The Supreme Court has recently granted certiorari regarding a case involving Section 230 and whether the federal law immunizes social media platforms when they make "targeted recommendations" of user-generated content—or only when the companies engage in "traditional editorial functions" regarding such content.⁹⁰⁵ As these cases proceed, the Supreme Court will provide vital clarity for U.S. policy makers attempting to develop constitutionally appropriate rules for news content shared over social media. As the new state laws further seek to articulate certain public obligations the platforms ought to fulfil in carrying on their operations, I provide a deeper analysis of how those public obligations might be achieved in the next section of this Part.

European Perspectives

Just as in the U.S., the advancement of communication technology has opened a new frontier in Europe where information may be more abundant than ever and its appropriate use more uncertain. But the characteristics of European societies, cultures, and history have shaped divergent perspectives on the nature of the issue itself and of appropriate responses. In the mid-1970s, Crozier described similar circumstances to those faced by Europe today: an overload of centralized bureaucratic decision-making systems brought on by an "explosion" of human interaction and communications.⁹⁰⁶ As discussed in the previous Part, early expectations about social media use for political expression and civic self-determination must now incorporate actualities like the French *groupes colère* protest

⁹⁰⁴ *Att'y Gen. of Fla. v. NetChoice, LLC*, No. ____, *Petition for a Writ of Certiorari* (21 September 2022) ("Petition"); see also *NetChoice, LLC v. Moody*, No. 21-12355 (11th Cir., 23 May 2022)

⁹⁰⁵ *Gonzalez v. Google, LLC*, No. 21-1333, *cert. granted* (03 October 2022)

⁹⁰⁶ Michel Crozier, 'Western Europe' in *The Crisis Of Democracy* by Michel Crozier, Samuel P Huntington, & Joji Watanuki (1975 NYU Press) 11, 11-12, 20-21

movements.⁹⁰⁷ For European citizens and decision-makers who prioritize an orderly, well-functioning society,⁹⁰⁸ such increases in public engagement, participation, and scrutiny brought on by social media (which might generally be considered laudable) has injected a challenging level of complexity to the process of rational governance. Crozier explains the expanding influence of European media in terms of its power to make certain events visible—that the reality of events depends on their being reported and disseminated, and so media creators such as journalists have “a structuring impact on public and social life.”⁹⁰⁹ While newspapers and television have varyingly retained certain regulatory mechanisms under law or through professionalism and audience fragmentation, social media seemingly vests anyone with the capability to make visible to everyone an endless variety of social issues, emotional appeals, and grievances. As a result, from the public’s perspective, bureaucratic mechanisms for implementing social goals appear increasingly unresponsive in light of rising issue visibility,⁹¹⁰ while governments increasingly view the citizenry as unmanageable and prone to manipulation and influence.⁹¹¹

European perspectives on fake news regulation accordingly seek to order competing social values against social costs. In the analysis of Helm and Nasu, for example, “the harmful consequences of fake news necessitate effective regulatory responses to reduce or eliminate its adverse impacts on the orderly society or political processes.”⁹¹² Regulation must, however, be made consistent with normative expectations of freedom of expression through consideration of legality, necessity, and proportionality.⁹¹³ The difficulty of defining what constitutes fake news complicates the determination of what content will foreseeably be subject to restriction—which has given rise to concerns about the German approach.⁹¹⁴ However, Helm and Nasu contend that “a legislative provision criminalising the creation or dissemination of fake news can be tailored in a way that specifically, and with sufficient precision, addresses the legitimate interests despite the inherent ambiguity of falsity as the constitutive element of the offence,” by incorporating safeguards “such as the requirement of malicious intent and the threshold of harm,” and the opportunity to contest allegations and raise affirmative defences of truth and good faith.⁹¹⁵ Such an approach necessarily entertains a culturally specific prioritization of certain values like public order relative to what Helm

⁹⁰⁷ See text accompanying Part I, ‘Social Media News Use’ (n 31) to (n 46)

⁹⁰⁸ Crozier (n 906) 45-46 (discussing core European political beliefs)

⁹⁰⁹ *ibid* 35

⁹¹⁰ *ibid* 35-37

⁹¹¹ *ibid* 41-42

⁹¹² Helm and Nasu (n 829) 305

⁹¹³ *ibid* 307-08

⁹¹⁴ *ibid* 309-10 (concerning the *NetzDG* holocaust denial and Nazi propaganda laws discussed above)

⁹¹⁵ *ibid* 324-25

and Nasu call “worship” of freedom of expression and a very European level of comfort with the social cost of “the chilling effect on socially beneficial free flow of information.”⁹¹⁶

The European Commission has emphasized its commitment to reaping the purported benefits of the EU digital single market (including “data-driven innovation” and “the potential to enhance citizens’ participation in society and democracy”) by creating a legal environment conducive to online intermediary platforms’ commercial growth.⁹¹⁷ The Commission specified that EU regulation should be “problem-driven” and not rely on traditional rules that “may impede online platform business models” (even though online platforms compete with/offer substitutes for traditional telecommunications and media services) and also should rely on “big data analytics” and “industry tools for ensuring application of legal requirements and appropriate monitoring mechanisms.”⁹¹⁸ In 2018, the EC set out its views on how Europe should address online disinformation, noting in particular the 2018 Facebook/Cambridge Analytica incident⁹¹⁹ and other threats to EU institutions ahead of the 2019 European Parliament elections.⁹²⁰ This European approach continues to stress “inclusive solutions,” while specifying expectations for social media platforms to meet, including the enlistment of “trusted flaggers” for information authentication and the provision of improved fact-checking capabilities. Likewise, according to the Commission, “Artificial intelligence, subject to appropriate human oversight, will be crucial for verifying, identifying, and tagging disinformation,” along with trustworthy electronic identification services.⁹²¹ As social media users are presumed to share content “without any prior verification,” media literacy initiatives and support for “high quality journalism” and “public media” are also emphasized.⁹²²

The 2018 Code of Practice on Disinformation⁹²³ put the Commission’s policies into action as a voluntary, self-regulatory regime, adopted initially by Facebook, Google, Twitter, Mozilla, various advertisers, and eventually Microsoft and later TikTok.⁹²⁴ The signatories committed to transparency around political advertising, combatting the misuse of automated systems with regard to their services (*i.e.*, ‘bots’), providing media literacy tools (*e.g.*, features

⁹¹⁶ Helm and Nasu (n 829) 326-27

⁹¹⁷ Commission, ‘Online Platforms and the Digital Single Market: Opportunities and Challenges for Europe’ COM (2016) 288 final, 25 May 2016

⁹¹⁸ *ibid* 5-6

⁹¹⁹ Discussed in Part I, ‘The Fake News Problem’

⁹²⁰ Commission, ‘Tackling online disinformation: a European Approach’ COM (2018) 236 final, 26 April 2018

⁹²¹ *ibid*

⁹²² *ibid*

⁹²³ EU Code of Practice on Disinformation (2018)

<<https://ec.europa.eu/newsroom/dae/redirection/document/87534>>

⁹²⁴ Directorate-General for Communications Networks, Content and Technology, ‘Shaping Europe’s digital future: 2018 Code of Practice on Disinformation’ (21 July 2022) <<https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation>> accessed 28 August 2022

that explain why users are seeing a particular advertisement), facilitating research and data sharing, tracking and evaluating disinformation, and partnering with the EC's network of fact-checkers.⁹²⁵ The Commission subsequently called for strengthening the Code of Practice, with particular regard to 'demonetizing' content producers identified as "purveyors of disinformation," expanding the coverage of fact-checking, and creating a permanent, EC-chaired taskforce to oversee implementation of the Code.⁹²⁶ The signatories adopted this revised Code in 2022.⁹²⁷ As discussed by Helm and Nasu, the Code aims for consistency with European normative considerations concerning freedom of expression and information, which demand sufficient justification for restrictions on the creation and dissemination of information—even when it is false.⁹²⁸

In the U.K., the 2018 Facebook/Cambridge Analytica incident prompted a more aggressive response from the Government, including extraordinary measures taken to investigate Facebook's decision-making processes through Parliamentary inquiries and document seizures.⁹²⁹ The U.K. Government subsequently produced a white paper in early 2019 detailing a plan for regulating social media platforms, "moving far beyond self-regulation" in order to address harmful disinformation that could "threaten public safety, undermine national security, fracture community cohesion and reduce trust."⁹³⁰ The white paper stressed however that the industry's own technology would be "part of the solution" and that online safety is a "shared responsibility."⁹³¹ The plan further called for an independent regulator to oversee and enforce the framework and for an online media literacy strategy to empower users.⁹³²

The Open Rights Group ("ORG") expressed concern over the purpose and scope of the Government's proposal, suggesting instead that policy intervention should focus on the business model of social media platforms and how they leverage user *data* rather than simply user-generated content.⁹³³ The ORG also criticized the Government's harms-based approach as unduly focused on social media users rather than the platforms and stressed

⁹²⁵ EU Code of Practice on Disinformation (n 923)

⁹²⁶ Commission, 'European Commission Guidance on Strengthening the Code of Practice on Disinformation' COM (2021) 262 final, 26 May 2021

⁹²⁷ The Strengthened Code of Practice on Disinformation 2022
<<https://ec.europa.eu/newsroom/dae/redirection/document/87585>>

⁹²⁸ Helm and Nasu (n 829) 307-308

⁹²⁹ Carole Cadwalladr, 'Parliament seizes cache of Facebook internal papers' (*The Guardian*, 24 November 2018) <<https://www.theguardian.com/technology/2018/nov/24/mps-seize-cache-facebook-internal-papers>> accessed 28 July 2022

⁹³⁰ *Online Harms White Paper* (n 836) 3, 22-25

⁹³¹ *ibid* 4

⁹³² *ibid* 9-10

⁹³³ Open Rights Group, 'ORG Policy Responses to Online Harms White Paper' (23 May 2019) <<https://www.openrightsgroup.org/publications/org-policy-responses-to-online-harms-white-paper/>> accessed 28 July 2022

protecting the rights of users to pursue free expression with one another.⁹³⁴ In addition, while ORG wrote supportively of co-regulation with an independent body (as opposed to state regulation or self-regulation),⁹³⁵ ORG strongly cautioned against approaches that would attach liability to platforms for third-party content, citing risks to fairness and due process as well as overbreadth resulting from imprecise definitions of actionable harms.⁹³⁶

After further consultation and development, U.K. Government's proposal was introduced as the Online Safety Bill in 2022.⁹³⁷ The Department for Digital, Culture, Media & Sport ("DCMS") heralded the bill as a "milestone," which aims to place existing U.K. telecommunications regulator Ofcom in charge of supervising social media platforms' compliance with the new framework.⁹³⁸ The framework creates duties of care for social media companies (registered 'user-to-user services' providers) in relation to regulated content.⁹³⁹ The subject content includes illegal content, like child pornography, and also lawful content that has been formally designated by the Secretary of State as harmful ('priority' content) or which nevertheless presents a "material risk of significant harm to an appreciable number of adults [and/or children] in the United Kingdom" ('non-designated' content).⁹⁴⁰ During the on-going amendment process, the Secretary of State issued a statement providing an indicative list of the kinds of content the government would seek to designate as priority content in secondary legislation (following consultation with Ofcom), recognizing that while the Government wants to keep the list responsive to "the most recent evidence and emerging harms," parliamentarians and other stakeholders also want their say in what harms the Government designates.⁹⁴¹ Various responsibilities regarding risk assessment and incident reporting attach where the platform is likely to be accessed by children, for instance, and 'Category 1' platforms of a certain threshold user-base bear specific responsibilities regarding journalistic content and "content of democratic importance."⁹⁴² The kinds of treatment of content listed by the Bill include taking down the

⁹³⁴ Open Rights Group (n 933) 2-3

⁹³⁵ *ibid* 3-4

⁹³⁶ *ibid* 4-5

⁹³⁷ Online Safety Bill, Bill 285 2021-22 (as introduced, 17 March 2022), Bill 121 2022-23 (as amended in Public Bill Committee, 28 June 2022) <<https://publications.parliament.uk/pa/bills/cbill/58-03/0121/220121.pdf>> ("Online Safety Bill")

⁹³⁸ Department for Digital, Culture, Media & Sport and Nadine Dorries, 'World-first online safety laws introduced in Parliament' (17 March 2022) <<https://www.gov.uk/government/news/world-first-online-safety-laws-introduced-in-parliament>> accessed 28 August 2022

⁹³⁹ Online Safety Bill (n 937) Part 3 Chapter 2

⁹⁴⁰ *ibid* Part 3 Chapter 7 s 49

⁹⁴¹ Department for Digital, Culture, Media & Sport and Nadine Dorries, 'Online Safety Update' Statement UIN HCWS194 (7 July 2022) <<https://questions-statements.parliament.uk/written-statements/detail/2022-07-07/hcws194>> accessed 28 August 2022

⁹⁴² Online Safety Bill (n 937) Part 3 Chapter 2

content, restricting user access, and altering the recommendation or promotion of the content.⁹⁴³

In relation to journalistic content, Category 1 platforms would be obligated to maintain ‘proportionate’ systems and processes in treatment of such content, as well as a dedicated and expedited complaints procedure for users facing action by the platforms with regard to content the users consider journalistic.⁹⁴⁴ Notably, the framework divides journalistic content into ‘news publisher content’ (created by a ‘recognized news publisher’ like the BBC or other U.K. registered entities meeting the conditions of the framework; and also other direct reproduction thereof, such as a re-tweet)⁹⁴⁵ and other ‘regulated user-generated content’ created for the purposes of journalism (provided the creator is within the U.K. and the content is directed towards or likely to be of interest to a significant number of U.K. users).⁹⁴⁶ A further amendment has been tabled that would require Category 1 platforms to ensure recognized news publishers’ content remains viewable and accessible even while undergoing moderation or review, and platforms would be obligated to notify recognized news publishers ahead of taking action and offer a right of appeal.⁹⁴⁷ In relation to content of democratic importance, the framework again stresses proportionality as the watchword in platform decisions to take action against a user, particularly in issuing warnings, suspensions or bans. The framework likewise distinguishes news publisher content and regulated user-generated content that “is or appears to be specifically intended to contribute to democratic political debate” in the U.K.⁹⁴⁸

Other key features of the Online Safety Bill include its provisions for Category 1 social media platforms to carry out impact assessments regarding users’ freedom of expression when deciding on safety measures and policies.⁹⁴⁹ The Bill would also require Category 1 platforms to make available to users tools for filtering out certain content and/or non-verified users⁹⁵⁰ and also services for user identity verification.⁹⁵¹ ‘Digital identities’ represent another plank of the broader U.K. Government strategy for public confidence online which has undergone extensive development in the past few years.⁹⁵² Similarly, the

⁹⁴³ Online Safety Bill (n 937) Part 3 Chapter 2 s 13

⁹⁴⁴ *ibid* Part 3 Chapter 2 s 16

⁹⁴⁵ *ibid* Part 3 Chapter 7 s 49 & s 50

⁹⁴⁶ *ibid* Part 3 Chapter 2 s 16

⁹⁴⁷ Department for Digital, Culture, Media & Sport and Nadine Dorries, ‘Tougher protections for journalism added to online safety laws’ (07 July 2022) <<https://www.gov.uk/government/news/tougher-protections-for-journalism-added-to-online-safety-laws>> accessed 28 July 2022

⁹⁴⁸ Online Safety Bill (n 937) Part 3 Chapter 2 s 15

⁹⁴⁹ *ibid* Part 3 Chapter 2 s 19

⁹⁵⁰ *ibid* Part 3 Chapter 2 s 14

⁹⁵¹ *ibid* Part 4 Chapter 1 s 57

⁹⁵² Department for Digital, Culture, Media & Sport and Matt Warman, ‘Government sets out new plans to help build trust in use of digital identities’ (11 February 2021)

Government has moved forward with online media literacy interventions aimed at building “audience resilience to dis- and misinformation,”⁹⁵³ in line with the determinations set forth in its 2019 white paper. In its capacity as the relevant regulator, Ofcom would be empowered to require annual transparency reports from service providers⁹⁵⁴ which could include information about harmful content on the platforms, terms of service violations, reporting systems and other processes.⁹⁵⁵ The Online Safety Bill further directs Ofcom to establish an advisory committee on disinformation and misinformation, consisting of representatives of the platforms and their users and also experts “in the prevention and handling of disinformation and misinformation online” who would advise Ofcom with regard to subject-relevant information gleaned from the annual transparency reports.⁹⁵⁶

Like the U.S., Europe also experienced the expanded influence of communications technology in the decades after World War II. But the European experience was mediated through substantially different political structures, diverse cultures, and long histories. The advances of the Internet age have made those factors no less complex. For countries like Germany, clear priorities about holocaust denial and Nazi propaganda underpin specific criminal sanction for those uses of social media. More widely, the E.U. pursues self-regulation by social media companies—relying on the firms to shoulder the administrative burden themselves—while providing support for society through transparency requirements and media literacy initiatives and the like. The U.K., however, perhaps more confident of its professional bureaucrats, aims to adopt a co-regulatory approach that may serve as the model for European fake news regulation going forward, provided it proves both efficacious and consistent with normative expectations. The Online Safety Bill presents an opportunity to assess how effectively regulatory power can be exercised by an agency like Ofcom in combination with the social media companies. In Part Five, I examine the prospects for its new regulations in relation to media regulation generally and more specifically to the platform capabilities discussed later in this Part.

<<https://www.gov.uk/government/news/government-sets-out-new-plans-to-help-build-trust-in-use-of-digital-identities>> accessed 28 July 2022; Department for Digital, Culture, Media & Sport and Julia Lopez, ‘UK digital identity and attributes trust framework – beta version’ (13 June 2022)

<<https://www.gov.uk/government/publications/uk-digital-identity-and-attributes-trust-framework-beta-version>> accessed 28 August 2022

⁹⁵³ Department for Digital, Culture, Media & Sport, ‘Media Literacy Taskforce Fund’ (18 July 2022)

<<https://www.gov.uk/guidance/media-literacy-taskforce-fund>> accessed 28 July 2022

⁹⁵⁴ Online Safety Bill (n 937) Part 4 Chapter 3 s 64

⁹⁵⁵ *ibid* Schedule 8 Part 1

⁹⁵⁶ *ibid* Part 7 Chapter 7 s 131

Trans-Atlantic Perspectives

Specific concern with election integrity animates a third perspective relative to the American debates on public interest rationales and the U.K. public safety rationale. Intelligence and security services on both sides of the Atlantic view elections-related fake news as a security *threat*. In 2021, the U.S. National Intelligence Council (a joint body comprised of U.S. federal law enforcement, surveillance, and other agencies) released a declassified report on its assessment of foreign threats to the 2020 U.S. federal elections, in which social media figured heavily.⁹⁵⁷ This Intelligence Community Assessment stated that “the growth of internet and social media use means foreign actors are more able to reach US audiences directly, while the tools for doing so are becoming more accessible.”⁹⁵⁸ While the declassified report did not provide any supporting information or evidence on which the security services relied, the report claimed that foreign adversaries had used online “influence narratives” “to amplify mistrust in the electoral process” and that “proactive information sharing with social media companies facilitated the expeditious review, and in many cases removal, of social media accounts covertly operated by Russia and Iran.”⁹⁵⁹ Platforms like Twitter have explicitly stated that they have taken action against Iranian user accounts “based on intel provided by the @FBI” (the Federal Bureau of Investigation, a U.S. federal law enforcement agency).⁹⁶⁰ The report also suggests non-state actors like foreign cybercriminals and “hacktivists” may target elections or campaigns for reasons of politics or profit.⁹⁶¹

In 2022, the U.S. Department of Homeland Security (“DHS”) assembled a working group called the Disinformation Governance Board, which DHS Secretary Mayorkas described as follows:

“What it will do is gather together best practices in addressing the threat of disinformation from foreign state adversaries, from the cartels, and

⁹⁵⁷ National Intelligence Council, ‘Foreign Threats to the 2020 US Federal Elections’ (10 March 2021) Intelligence Community Assessment 2020-00078D
<<https://www.dni.gov/files/ODNI/documents/assessments/ICA-declass-16MAR21.pdf>> accessed 28 July 2022 (hereinafter ‘Intelligence Community Assessment’)

⁹⁵⁸ *ibid* 1

⁹⁵⁹ *ibid* 1-2

⁹⁶⁰ Twitter Safety (01 October 2020)

<<https://twitter.com/TwitterSafety/status/1311462538056544258?s=20>> (accessed 28 August 2020)

⁹⁶¹ Intelligence Community Assessment (n 957) 8-9

disseminate those best practices to the operators that have been executing in addressing this threat for years.”⁹⁶²

While Mayorkas said the focus of the Board would be on “entities that represent a threat to ‘the security of the homeland,’” the announcement of the Board proved controversial, and journalists from liberal and conservative media outlets alike noted the ominous, heavy-handed tone of the venture.⁹⁶³ Following the public response and a review by the Homeland Security Advisory Council, the Board was terminated several months later.⁹⁶⁴ Yet the Advisory Council report nevertheless called for the DHS to “be able to address disinformation threat streams that can undermine the security of our homeland” and recommended re-prioritizing existing components of the Department to assess whether “publicly disseminated disinformation impedes missions assigned to the agency by law” and to disseminate “correct information” to the public.⁹⁶⁵

In addition, the Advisory Council report specifically pointed at foreign actors using “social media platforms to create and amplify narratives that further their ends” and stated that the DHS “has a responsibility to protect our national security by warning of these activities and sharing this information with other relevant government agencies.”⁹⁶⁶ The DHS’s Cybersecurity and Infrastructure Security Agency (“CISA”) further describes risks to election security from disinformation and “false statements” as relevant to its core missions.⁹⁶⁷ Among the response measures taken by the CISA are what it calls “products,” including a “Foreign Interference Taxonomy;” a graphic novel, presumably for the benefit of security-minded comic book readers, that “demonstrates how threat actors capitalize on political and social issues (especially around election cycles) to stealthily plant doubt in the minds of targeted audiences and steer their opinion”; and a “Mis-, Dis-, and Malinformation Planning and Incident Response Guide for Election Officials.”⁹⁶⁸ As discussed in the previous Part, the decisions baked into definitions and taxonomies of fake news circumscribe both the ‘problem’ itself and the kinds of solutions that logically follow after.

⁹⁶² Seth Smalley, ‘What we know about the DHS’ Disinformation Governance Board’ (*Poynter Institute*, 05 May 2022) <<https://www.poynter.org/reporting-editing/2022/what-is-department-homeland-security-disinformation-governance-board/>> accessed 15 November 2022

⁹⁶³ *ibid*

⁹⁶⁴ Department of Homeland Security, ‘Following HSAC Recommendations, DHS terminates Disinformation Governance Board’ (24 August 2022) <<https://www.dhs.gov/news/2022/08/24/following-hsac-recommendation-dhs-terminates-disinformation-governance-board>> accessed 15 November 2022

⁹⁶⁵ Homeland Security Advisory Council, Disinformation Best Practices and Safeguards Subcommittee, *Final Report* <https://www.dhs.gov/sites/default/files/2022-08/22_0824_ope_hsac-disinformation-subcommittee-final-report-08242022.pdf> accessed 15 November 2022, 12-14

⁹⁶⁶ *ibid* 14

⁹⁶⁷ *ibid* 17

⁹⁶⁸ *ibid* 18-20

In the European theatre, additional attention to “foreign interference” in elections, particularly through social media, has proceeded from the 2018 Paris Call for Trust and Security in Cyberspace, to which numerous governments, civil society, and industry groups have signed on.⁹⁶⁹ Sweden, Finland, and Latvia have all applied a “total defense” strategy to “educate and protect” their populations from “false information and cyber-attacks as well as many other threats.”⁹⁷⁰ The language of threat and security has notably been employed by European Commission President von der Leyen, who stated that “our democratic systems and institutions have come increasingly under attack in recent years from those who wish to divide or destabilize our Union.”⁹⁷¹ Perhaps owing to the prerequisite that European interference with human rights must be clearly prescribed by law, recommendations by strategists for European policymakers have featured a more nuanced approach than in the U.S., with greater attention to how an instance of “interference” should be defined.⁹⁷²

Accordingly, the German Marshall Fund’s Alliance for Securing Democracy (“ASD”) project highlighted to European lawmakers the challenges faced by social media companies in developing consistent election integrity policies (which are discussed in the next section of this Part), contending that “the lack of a standard definition for the concept of interference has led private companies to set inconsistent policies in their terms of service when policing online interference, including disinformation.”⁹⁷³ The ASD contrasted the European Council’s 2019 *pro forma* pronouncement on “disinformation and manipulative interference” threats with American policy language, which emphasizes determination of malignant intent; as well as the Australian approach, which focuses on the transparency of foreign influence activities.⁹⁷⁴ But the ASD also considered that “framing a definition solely around legality is challenging for several reasons”:

“First, legal frameworks often take time to catch up to the activities in question. . . . The new European Commission will only take up the question of regulation of industry to prevent against disinformation . . . in the coming years. Waiting for a determination that a certain kind of technological attack is illegal before calling it interference would be overly limiting for policymakers trying to respond to it. Second, legal frameworks often rely on existing laws that do not take into account the unique circumstances of

⁹⁶⁹ Kristine Berzina and Etienne Soula, ‘Conceptualizing Foreign Interference in Europe’ (*Alliance for Securing Democracy*, 19 March 2020) <<https://securingdemocracy.gmfus.org/wp-content/uploads/2020/03/Conceptualizing-Foreign-Interference-in-Europe.pdf>> accessed 28 July 2022

⁹⁷⁰ *ibid* 1

⁹⁷¹ *ibid* 2

⁹⁷² *ibid* 2-3

⁹⁷³ *ibid* 3

⁹⁷⁴ *ibid* 3-5

interference operations. . . . The determination of what is ‘illegal content’ is based on existing laws that define hatred, depictions of violence, and child pornography, defamation, and forgery. A major challenge of this approach is that foreign disinformation or manipulated content ‘is rarely so overtly inflammatory as to be considered criminal.’”⁹⁷⁵

This reasoning ultimately leads the ASD strategists to urge the EU to adopt an approach that focuses on intent and transparency instead of a particularized unlawful act of ‘interference’ in order to maintain flexibility of response for government agencies and private industry.⁹⁷⁶

Following the 2020 U.S. elections, the Government of Canada teamed with ASD and Microsoft to produce a ‘multi-stakeholder’ report on election interference threats which also discussed the challenges of acting permissibly in relation to political definitions of interference.⁹⁷⁷ Their discussion similarly focused on countering disinformation campaigns against democracies that “attempt to interfere with elections by sowing doubt in electoral processes, aggravating societal polarization, and applying a host of other nefarious tactics.”⁹⁷⁸ The multi-stakeholder report also emphasized cooperation with professional journalists and social media platforms⁹⁷⁹ and lauded the increased involvement of U.S. federal authorities (like CISA) and civil society partners in the recent U.S. elections.⁹⁸⁰

“Citizens need access to diverse and reliable sources of information in order to develop informed views, make good decisions and exercise sound judgement. Unfortunately, the ever-growing problem of widespread disinformation and toxic online commentary found in countless social media posts is eroding citizens’ faith in legitimate sources of information. Disseminated by both foreign and domestic actors, disinformation is, therefore, a threat to democracy itself, including elections.”⁹⁸¹

⁹⁷⁵ Berzina and Soula (n 1390) 8-9

⁹⁷⁶ *ibid* 9-11

⁹⁷⁷ Brad Smith, Karen Donfried, and Dominic LeBlanc, ‘Multi-Stakeholder Insights: A Compendium on Countering Election Interference’ (*Alliance for Securing Democracy*, 14 April 2021) <<https://www.canada.ca/content/dam/di-id/documents/compendium-eng.pdf>> accessed 28 July 2022, 10-13

⁹⁷⁸ *ibid* 20

⁹⁷⁹ *ibid* 21-23

⁹⁸⁰ *ibid* 39

⁹⁸¹ *ibid* 40

Its recommendations included additional partnerships with industry and various communities, digital literacy programs “to help citizens recognize disinformation,” and support for trusted fact-checkers as part of the “information ecosystem.”⁹⁸²

Public policy approaches to the issue of fake news reflect a wide array of historical, constitutional, and political exigencies specific to the U.S., U.K., and elsewhere. As covered in this section, some U.S. lawmakers increasingly view the prevailing ‘hands-off’ approach to technology firms as untenable, but vary on specifics. Differing notions of public interest, liberty, transparency, and accountability inform various federal regulatory proposals and at least two state law proposals so far, but any regulation in the U.S. must reckon with existing laws like Section 230 and the Constitutional protections of the First Amendment. European lawmakers similarly stress transparency and accountability, but so far have relied on voluntary agreements and self-regulation, owing to an array of complex and competing socio-cultural factors. The U.K. has moved towards a more intensive, public safety-based rationale that would position Ofcom as an independent regulator of social media platforms. And on both sides of the Atlantic, security organizations push for greater authority to respond to fake news as a security threat, particularly in the elections context.

But across each of these perspectives, the social media companies themselves retain a central role in achieving the aims of government policymakers—whether through co-regulation, cooperation, or more extensive partnerships. Co-regulation alongside independent regulators, the European approach, necessitates the platforms do the work of content moderation, albeit in a fashion that meets the responsibilities and guidelines set forth by the regulator. Even in instances like the Florida law’s proposed liability for a platform’s inconsistent application of moderation standards, the social media companies are still expected to undertake activities to evaluate the content their users generate. Where initiatives demand a more aggressive stance by social media companies against supposedly dangerous news content, the policies and technical capabilities of the platforms become instrumental to achieving policymaker’s objectives.

Analysing Allocation Decisions

In this section, I consider what it means for us to make particular decisions about how social media should function in our socio-cultural body. These decisions take the form of accountability determinations regarding social media news content—who will be held responsible by whom. These determinations prioritize particular interests and values, as discussed in the preceding section, be they commercial, political, or individual. The likely

⁹⁸² Smith, Donfried, and LeBlanc (n 977) 40-42

impact of these decisions on free speech and the press can be projected based on the structural factors that characterize our institutions. The institutions I consider in this section include the social media platforms, regulatory or administrative state agencies, the judiciary, and legislatures. Each of these institutions can express authority over social media news content, each represents particular interests, and each has its own history in shaping the function of media systems. This Part concludes with a discussion of the trade-offs identified in this section.

Status Quo: Social Media Platform Control

Under the current allocation of regulatory authority, the social media platforms hold responsibility for news content carried on their services. As detailed in Part One, social media news use has been recognized as both positive and negative, eliciting significant concerns over its role in civic discourse and elections.⁹⁸³ The platforms are responsible for social media news content in terms of the architecture they provide for uploading and sharing information and the algorithmic controls they use for adjusting its reach and visibility.⁹⁸⁴ But the legal conventions under which social media platforms operate do not make the platforms accountable for the information users provide. More importantly, provisions like Section 230 shield the platforms from liability for what they *do* with that information. That means they don't have to play fair when it comes to favouring one source over another or picking and choosing which narratives achieve currency on their services.

This situation features a novel legal convention but a historically familiar arrangement of regulatory force. As described in Part Two, the U.S. allowed for the telegraph industry to develop without restriction of the industry's control over what messages were transmitted by whom. As a result, Western Union achieved a monopoly over telegram distribution and entered into a collusive relationship with the Associated Press that excluded its rivals and circumscribed public discourse. Anti-monopoly reformers criticized the collusion between Western Union and the Associated Press as "especially dangerous in a democracy" that depended on public intelligence as a check on government.⁹⁸⁵ The U.S. House of Representatives and the Senate both highlighted the abuses by the two companies in the 1870s, documenting "instances in which Western Union cut off transmission of news reports to papers that criticized the telegraph company and the content of AP dispatches."⁹⁸⁶ Inkling of this relationship between content distributor and content provider can be seen in the "partnerships" between platforms like Facebook and established news media as

⁹⁸³ Part I, 'The Fake News Problem'

⁹⁸⁴ Part II, 'Platform Policies and Technical Capabilities'

⁹⁸⁵ Daniel J Czitrom, *Media and the American Mind: From Morse to McLuhan* (Univ N Carolina P 2010) 25-26

⁹⁸⁶ *ibid* 26

described in Part Two.⁹⁸⁷ Facebook's mass media partners have strong financial incentives to exclude news rivals and upstarts and pressure Facebook to direct traffic towards their own outlets. Independent media outlets' reports of precipitous drops in web traffic after platform policy changes are highly suggestive of this eventuality.⁹⁸⁸

The platforms also have their own incentives under the status quo. The current business model of the platforms relies on attracting audiences to market to advertisers.⁹⁸⁹ The platforms rely on the vast amounts of demographic and behavioural data they generate about their users to provide added value to advertisers, and the flow of that data depends on keeping users online and engaged through services like content sharing. The platforms have identified divisive content as especially engaging, and likely make sophisticated determinations about the amount and character of such content that users will respond to.⁹⁹⁰ A loss of authority to make those determinations, even to a partial extent, in turn represents a potential loss of opportunities to maintain levels of marketable audience engagement. The threat of public policy that would change the control social media platforms enjoy over content decisions thus also incentivizes the platforms. As seen in the changes made by the major platforms in the wake of public scrutiny following the Cambridge Analytica story, the platforms do respond to threats of regulation, and appear willing to accede to specific government requests for internal moderation policy changes in exchange for retained authority.⁹⁹¹ In addition, because regulatory power is currently allocated to the platforms, the usual incentive for industry to 'capture' its public regulator is inverted: rather than former industry insiders moving into government roles, former government officials,⁹⁹² particularly security service and intelligence agents,⁹⁹³ take on management and operational roles at the social media companies.

Another consequence of status quo content moderation arrangements is the continued elusiveness of social media transparency. To the extent that social media

⁹⁸⁷ Part II, 'Platform Policies and Technical Capabilities' – Facebook

⁹⁸⁸ Matt Taibbi, 'Who Will Fix Facebook?' (*Rolling Stone*, 26 Nov 2018)

<<https://www.rollingstone.com/politics/politics-features/who-will-fix-facebook-759916/>> accessed 27 July 2022

⁹⁸⁹ See, e.g., Aleksandra Gabicka and Andreas Heinemann, 'Social Media Competition Law' (2014) 37(2) *World Competition L and Econ Rev* 149, 154-56 (describing how platforms like Facebook operate as 'two-sided' markets)

⁹⁹⁰ Part I, 'Social Media News Use' and Part II, 'Platform Policies and Technical Capabilities'

⁹⁹¹ Part I, 'The Fake News Problem'

⁹⁹² Former Deputy PM Nick Clegg, for example, serves as Facebook-parent company Meta's "president of global affairs," focusing on regulatory issues; *BBC*, 'Nick Clegg gets bigger role at Facebook owner' (*BBC.com*, 16 Feb 2022) <<https://www.bbc.com/news/uk-60410636>> accessed 01 April 2023

⁹⁹³ e.g., Alan MacLeod, 'Meet the ex-CIA Agents Deciding Facebook's Content Policy' (*MintPress News*, 12 July 2022) <<https://www.mintpressnews.com/meet-ex-cia-agents-deciding-facebook-content-policy/281307/>> accessed 01 April 2023 (detailing politically sensitive senior roles at Facebook now held by former security service personnel)

companies rely on large datasets to automate binary content classifications (identifying user content as news or not, harmful or not, fake or not), the platforms already face substantial difficulty in articulating what those binary classifications are actually *like*. A company like Twitter, for instance, could tell users that a particular tweet has been labelled as misinformation, but to be transparent about whether the company considers a user's tweet to be misinformation requires more than just a policy statement defining 'misinformation.' Parts of Twitter's policy definition do contain bright line measures which can be easily explained to users, such as when a user posts a claim that has been "confirmed to be false by external, subject-matter experts." Users can then take or leave such an explanation based on their assessment of the expert opinion. But other elements of Twitter's policy definition cannot be so easily interrogated by users. To explain whether purported misinformation has been "shared in a deceptive or confusing matter," Twitter would have to articulate what characteristics of user sharing behaviour, source, audience, or content signals triggered the identification of the content as misleading. Likewise, explaining the "level of potential harm" associated with a tweet (which triggers various treatments of the content like labelling or removal) would require Twitter to articulate how the company connects certain signals about the content to a potential harm (such as undermining faith in election processes).

A transparent explanation for such a content moderation decision might look like this:

Your tweet (see attached) has been identified as misinformation based on the following factors, which our automated system identified as being more or less similar to our model for what deceptive or confusing sharing looks like—

1. negative tone (86% similar)
2. simplistic writing style (42% similar)
3. sharing among 'vulnerable groups' (62% similar)
4. expert ratings of the source you linked (90% similar)

Your tweet has an unweighted average score of 70% similarity to what we categorize as misinformation. We have also scored the level of potential harm associated with the semantic content of your tweet (identified by simple word search comparison with our database)—

1. news topic (low potential harm, scored 1)
2. elections topic (high potential harm, scored 3)
3. public figure topic (medium potential harm, scored 2)

Your tweet scored an aggregate risk of potential harm calculated as:

$$0.7 * (1 + 3 + 2) = 4.2$$

Risk scores between 3 and 5 receive warning labels; risk scores above 5 are subject to removal from our service. Accordingly, we have added a label to your tweet identifying it as misinformation.

In reality, the models used by the social media companies likely incorporate more than just four factors in classifying content, but the four shown here reflect some of the core factors examined in the machine learning studies discussed in Part Two. A truly transparent explanation of content moderation decisions would require companies to specify all of the factors they use in classifying content, as well as the statistical weights they assign to rank the relative importance of each factor. Additionally, the companies would need to explain which potential harm (like offline violence or undermined faith in elections processes) is associated with what semantic content and how the platforms have ranked the relative seriousness of particular harms. These explanations would not be merely technical but also include the legal and policy judgements made by the social media companies.

Social media companies currently do not provide this level of detail for a number of reasons. First, such disclosures would make more explicit the kinds and amount of data that the platforms routinely collect from and about their users. While a platform may be leery about revealing this to users, a more pressing business reason for a platform to keep such information secret would be to avoid disclosing technical capabilities to its competitors. Likewise, details about the operation of a platform's content models would likely be of keen interest to rival platforms, and effective models would be easier to reverse-engineer with the information released in transparent disclosures. Furthermore, a platform may be unable to disclose factors related to input received from third parties like fact-checkers or subject-matter experts and government partners. The private terms of a platform's contractual relationships with such third parties are generally unavailable for public review, but the recent 'Twitter Files' exposés have shown some of the high-level coordination between Twitter and "industry peers" with the U.S. government's security services, including the sharing of classified information⁹⁹⁴ as well as the advertising by the security services of specific user accounts for "additional action."⁹⁹⁵ Where inputs from these sources touch on

⁹⁹⁴ Matt Taibbi, 'Twitter Files' 24 (16 December 2022)

<<https://twitter.com/mtaibbi/status/1603857632091148288?s=20&t=2OrkAnvOz9O0QSMHh-ow4A>>
accessed 04 January 2023

⁹⁹⁵ Matt Taibbi, 'Twitter Files' 18 (16 December 2022)

<<https://twitter.com/mtaibbi/status/1603857603867598859?s=20&t=2OrkAnvOz9O0QSMHh-ow4A>>
accessed 04 January 2023

contractual obligations or national security restrictions, the platforms may be legally prohibited from disclosing sufficient useful information about moderation decisions.

Under the regulatory approaches proposed in the American States of Florida and Texas, the platforms would, however, face some obligations to disclose information along these very lines. As discussed in the previous section, Florida's SB 7072 would require regulated platforms to provide users with a "thorough explanation" of enforcement actions and make routine disclosures about algorithmic content moderation and detail their moderation standards.⁹⁹⁶ Texas's HB 20 requires disclosures about moderation, ranking algorithms, and user-data usage, as well as an explanation to users concerning content removals.⁹⁹⁷ Yet if such laws go into force, the platforms would still face the same pressures they do now to disclose as little as possible for business competition, relational, and legal reasons. These transparency requirements would be at odds with those interests of the social media companies and their partners. Unsurprisingly, the transparency requirements of each state law have been key parts of the court challenges by the social media industry against Florida and Texas. But regardless of the ultimate fate SB 7072 or HB 20, any transparency requirements imposed on the platforms that conflict with these interests will likely face resistance.

Where then do these status quo relationships and incentives leave the users of social media? Despite generating vast amounts of data for the platforms' recommendation engines and the attraction for advertisers, the users of the social media platforms are very poorly structurally positioned in the social media system to influence the platforms' accountability determinations. One factor is the difficulty of transparency on the part of the platforms in moderation decisions, as discussed above. Users will continue to have a hard time getting enough information to evaluate the quality of platform accountability determinations. A further factor is the relational imbalance between users and platforms embodied by the take-it-or-leave-it nature of 'terms of service' or 'terms of use' agreements.⁹⁹⁸ These "clickwrap" contracts do not afford users the opportunity for arms-length negotiation with the platforms over content, privacy, or monetization preferences, even assuming users take the time to read the contracts before assenting to the terms.⁹⁹⁹ Waulters, Lievens, and Valcke have outlined a number of ways the relationship between users and platforms could be improved,

⁹⁹⁶ SB 7072 §§ 501.2041 (2), (3)

⁹⁹⁷ HB 20 §§ 120.051, 120.053

⁹⁹⁸ Allen P Grunes and Maurice E Stucke, 'No Mistake About It: The Important Role of Antitrust in the Era of Big Data' (*Antitrust Source*, April 2015) University of Tennessee Legal Studies Research Paper No 269, 6-7 <<https://ssrn.com/abstract=2600051>> accessed 01 April 2023 (marking the contractual relationship as an important difference between social media and traditional media)

⁹⁹⁹ Jonathan Obar and Anne Oeldorf-Hirsch, 'The Clickwrap: A Political Economic Mechanism for Manufacturing Consent on Social Media' (July–Sept 2018) *Social Media + Society* 1, 2-6

including public obligations to provide better terms and technical solutions like interactive, customisable contracts.¹⁰⁰⁰ But the platforms have little incentive to vary their current practices, and users little resources to articulate what they might want from the platforms.

In 2019, Facebook co-founder Chris Hughes publicly called for his former company to be broken up, arguing that its dominant position in the market makes it a threat to American democracy.¹⁰⁰¹ But forcing Facebook to spin off aspects of its business (like Instagram) would do nothing to change the structural factors of social media—the two-sided business model, closed and centralised processes, and its relationship to the users—that most impact accountability for content. Though Hughes cites defeats of Standard Oil and the telephone monopoly as past victories of antitrust crusaders,¹⁰⁰² the successors of those businesses are now once again under only two or three umbrellas instead of one.¹⁰⁰³ Like the oil industry and the telephone industry, returns on scale for platforms like Google and Facebook result from intensive capital investment in technical infrastructure that allows them to offer a ‘free’ product to users: more servers, faster connectivity, distinctive intellectual property, and better programmers.¹⁰⁰⁴ Additional competitors in the social media market will have the same incentives to grow large, while the ‘network effects’ of larger platforms will continue to draw in (and lock in) users who want to be able to connect to more and more users.¹⁰⁰⁵ Rahman points to these infrastructural characteristics as the appropriate level of functional analysis,¹⁰⁰⁶ and in the case of the telephone companies, their treatment as public utilities has yielded more effective public intervention than has attention to their market power.¹⁰⁰⁷

Unlike the early Internet moderators who could exercise personal judgment (essentially handicraft in the sense of Mumford¹⁰⁰⁸) over whether a participant was a valuable addition to small networks like a chatroom or bulletin board,¹⁰⁰⁹ the mechanical

¹⁰⁰⁰ Ellen Wauters, Eva Lievens, and Peggy Valcke, ‘Social Network Sites’ Terms of Use: Addressing Imbalances in the User-Provider Relationship through Ex Ante and Ex Post Mechanisms’ (2014) JIPITEC 139

¹⁰⁰¹ Chris Hughes, ‘Opinion: It’s Time to Break Up Facebook’ (*New York Times*, 09 May 2019) <<https://www.nytimes.com/2019/05/09/opinion/sunday/chris-hughes-facebook-zuckerberg.html>> accessed 28 July 2022

¹⁰⁰² *ibid*

¹⁰⁰³ Standard Oil’s businesses are now primarily held by ExxonMobil, Chevron, and BP; AT&T has reacquired most of the ‘Baby Bells’, alongside Bell Atlantic successor Verizon

¹⁰⁰⁴ See, e.g., Matthew Hindman, *The Internet Trap: How the Digital Economy Builds Monopolies and Undermines Democracy* (Princeton UP 2018) (describing supply-side returns on scale as central to the commercial Internet)

¹⁰⁰⁵ Cory Doctorow, ‘Right or Left, You Should Be Worried About Big Tech Censorship’ (*Electronic Frontier Foundation*, 16 July 2021) <<https://www.eff.org/deeplinks/2021/07/right-or-left-you-should-be-worried-about-big-tech-censorship>> accessed 28 July 2022

¹⁰⁰⁶ K Sabeel Rahman, ‘The New Utilities: Private Power, Social Infrastructure, and the Revival of the Public Utility Concept’ (2018) 39 *Cardozo Law Review* 1621, 1668-69 and 1682

¹⁰⁰⁷ Discussed *infra*, – Legislative Control

¹⁰⁰⁸ Lewis Mumford, *Technics and Civilization* (7th edn, Routledge & Kegan Paul 1955) 277, 350-51

¹⁰⁰⁹ Discussed in Part II, ‘Platform Policies and Technical Capabilities’

operation of algorithmic moderation at scale points towards the sameness and uniformity that characterize mass production. The interplay between network logic and algorithmic homogenization further illuminates observations about the rise of “alternative” social media sites, where a study by Pew found that 15% of prominent accounts on seven newer social media networks had been banned or demonetized on mainstream platforms like Twitter and Facebook.¹⁰¹⁰

While only accounting for a small percentage of news consumers in 2022,¹⁰¹¹ the rise of alternative platforms shows that for some news consumers, the value of (purportedly) less restrictive networks rivals that of more-intensively moderated networks. Additionally, that value offsets or exceeds the costs of switching from mainstream platforms or maintaining accounts on multiple networks—although for users who have been banned from a mainstream platform, switching costs essentially fall to zero. Pew also noted that news consumers on alternative and mainstream social media platforms tend to differ substantially in political affiliation, with those who get news from at least one of the seven alternative platforms studied leaning Republican¹⁰¹² and news consumers on established platforms leaning Democrat.¹⁰¹³ The Pew study additionally found that news consumers on alternative social media sites preferred the protection of freedom of information over content restriction by private platforms, even if it meant that false information could be published.¹⁰¹⁴

Laws like Section 230 leave power over content in the hands of the platforms—which remain vulnerable to dysfunction and misdirection amidst a host of incentives that do not align with users’ interests in access to information and discussion. Accordingly, we should expect that social media platforms will continue current moderation practices or even intensify them with regard to political news content in particular. This will result in continued and perhaps worsening potential for social media dysfunction in relation to broader society, particularly if users leave ‘broad church’ established networks like Facebook

¹⁰¹⁰ Galen Stocking and others, ‘The Role of Alternative Social Media in the News and Information Environment’ (Pew Research Center, 06 October 2022) <<https://www.pewresearch.org/journalism/2022/10/06/the-role-of-alternative-social-media-in-the-news-and-information-environment/>> accessed 25 April 2023

¹⁰¹¹ *ibid* (6%)

¹⁰¹² *ibid* (averaging 66% Republican or Republican-leaning)

¹⁰¹³ Stocking and others (n 1010) (averaging 55% Democrat or Democrat-leaning); Pew Research Center, ‘Social Media and News Fact Sheet’ (20 September 2022) <<https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/>> accessed 28 April 2023 (Twitter is higher than the average, with 65% of Twitter news users Democrat or Democrat-leaning, while Facebook news users split more evenly 43%-51% Republican/Democrat)

¹⁰¹⁴ Pew Research Center (n 984) (64% of alternative news consumers in favor, compared to 37% among U.S. adults broadly)

to seek more explicitly (and narrowly) partisan ones.¹⁰¹⁵ My experiment demonstrated that users' behavioural data incorporated into platform moderation processes can reflect drives to prioritize group identity over discernment of fact. Other studies have shown that even finely-tuned algorithms have inherent difficulties in stopping fake news. If platforms like Twitter and Telegram increasingly become the political 'echo-chambers' pundits fear, the platforms' own algorithms will have had a hand in delivering networks that prioritize sameness rather than difference as one group or another reaches a critical mass.¹⁰¹⁶ Those networks will lose the advantage of discursive diversity that can make large networks socially functional (as a public sphere) in the first place.¹⁰¹⁷

Coming Soon: Regulatory Agency Control

An alternative allocation of responsibility for fake news content could vest regulatory authority in the administrative state—the kinds of government agencies that handle licensure and regulatory compliance monitoring in other technical industries such as broadcasting. Under this allocation, a regulatory agency like Ofcom decides what content to hold the platforms accountable for distributing based on public guidelines. Relying on a regulator to manage responsibility for social media content reflects the core ethos of regulatory agencies generally: regulators can bring to bear the same kind of specialization and expertise as managers can in the private sector, but with added responsiveness to the public. But as discussed in the previous section, regulatory agencies nevertheless may tend to marginalize public participation in regulation, owing to that same specialization that lets the agencies interface with complex industries.

The leading example of deciding to allocate authority to a regulatory agency is the U.K.'s Online Safety Bill.¹⁰¹⁸ What the Online Safety Bill aims to do is to make certain social

¹⁰¹⁵ By similar reasoning, this development could have considerable economic implications as well: fractured or disparate networks not only have less social value as *networks*, but also have less potential to realize economic returns on infrastructural scale discussed by Rahman (n 1006) 1670

¹⁰¹⁶ As discussed in Part IV, 'Implications', larger user groups' content may be advantaged over smaller groups' simply by the volume and frequency at which the dominant group can 'flag' rival content

¹⁰¹⁷ Cf. Mariano Croce and Andrea Salvatore, 'Why Does the Law Want Us to be Normal? Schmitt's Institutionalism and the Critique of the Liberal Legal Order' (Spring 2016) 93 *Cultural Critique* 32, 37-39 (suggesting, based on Schmitt's institutionalist-decisionist framework, that the social construction of normality requires just such a process of "ongoing exclusion and selective inclusion" as we see on social media platforms, and that the political selection of a means to order communities (e.g., social media networks) through institutions inevitably belies claims to being "proactively neutral, fair, and nondiscriminatory"—i.e., adding a legal gloss will not change the tendency towards less pluralistic social media that we see now, but will continue it under sovereign cover/excuse)

¹⁰¹⁸ Online Safety Bill, Bill 285 2021-22 (as introduced, 17 March 2022), Bill 121 2022-23 (as amended in Public Bill Committee, 28 June 2022) <<https://publications.parliament.uk/pa/bills/cbill/58-03/0121/220121.pdf>>; With the new regulations under the Digital Services Act coming into force, the E.U. has jumped ahead of the U.K. in pursuing a very similar framework; Regulation 2022/2065 of the European

media companies more responsive to external determinations about online content. To frame the Bill in terms of my allocation analysis, it aims to shift some control over content moderation away from the platforms. As detailed in the previous section,¹⁰¹⁹ the Bill obligates Category 1 platforms¹⁰²⁰ to moderate the content it specifies if a user so requests, restricts what actions they may take with regard to the content of ‘recognized news publishers,’ and subjects them to enforcement actions by Ofcom if they overstep the bounds of ‘proportionate systems and processes.’ The Bill thus prescribes what content for which the platforms will be accountable to Ofcom, and sets the terms by which Ofcom regulators will guide their decisions. The Bill itself has also changed¹⁰²¹ during its consideration: The removal from the Bill of future content designation updates through secondary legislation means that to a certain extent, the Parliament will forego exercising its power to respond to future public demands about making particular content accountable.

In obligating the platforms to moderate specified content, however, the Online Safety Bill continues to rely on the technical capabilities of the platforms to identify and make decisions about content. This co-regulatory approach carries with it several key risk factors that may worsen social media system dysfunction. These include unclear aims and tangled incentive structures as well as the limitations of the platforms covered previously. Being further obligated to do what they already do, the platforms will likely seek more data from users about content and could place more content under moderation than at present in order to comply with regulators.¹⁰²²

The problems facing regulation more broadly and co-regulation in particular are highlighted in a closer look at the Online Safety Bill as a law. Hadfield has argued for critical appraisal of laws by focusing on what a law *does*: how the law subjects “human conduct to the governance of rules.”¹⁰²³ Reed uses this approach as a basis for judging the quality of law “from a specifically legal perspective” by evaluating whether a law succeeds or fails at influencing behaviour to some useful end.¹⁰²⁴ In the particular context of cyberspace

Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act) (hereinafter “Digital Services Act Regulations”)

¹⁰¹⁹ Part V, ‘Public Policy and the Public Sphere’ – European Perspectives

¹⁰²⁰ Comparable to the Digital Services Act Regulations’ Article 33 ‘Very large online platforms’; the ‘Digital Services Coordinator’ of each E.U. member state is similarly situated to Ofcom in the U.K. as regulatory enforcer

¹⁰²¹ Compare Bill 285 2021-22 with Bill 220 2022-23 (as amended on re-committal, 19 December 2022)

¹⁰²² Corynne McSherry, Jillian C York, and Cindy Cohn, ‘Private Censorship is Not the Best Way to Fight Hate or Defend Democracy: Here Are Some Better Ideas’ (*Electronic Frontier Foundation*, 30 January 2018) <<https://www.eff.org/deeplinks/2018/01/private-censorship-not-best-way-fight-hate-or-defend-democracy-here-are-some>> accessed 26 July 2022

¹⁰²³ Gillian K Hadfield, *Rules for a Flat World*, ‘The Invention of Law’ (OUP 2016) 19-20

¹⁰²⁴ Chris Reed, ‘How to make bad law: Lessons from Cyberspace’ (2010) 73(6) *Modern Law Review* 903, 903-04

regulations, Reed identified a tendency for such law-systems to exhibit high levels of precision—a kind of certainty at the cost of complexity.¹⁰²⁵ Reed characterizes ‘precision’ law-making by reference to quantitative (rather than qualitative) measurements, a ‘tick-box’ approach to compliance, and obscured or non-obvious normative aims.¹⁰²⁶ Reed then showed how electronic signature and data protection laws achieved this precision but did not necessarily achieve their ostensible aims of promoting the use of digital signatures or restricting the processing of personal data.¹⁰²⁷ In the case of digital signature law, the complexity of compliance drove businesses away from using e-signatures. Data protection law, on the other hand, provided achievable compliance mechanisms that did little to protect data. Per Reed, such failures reflected the low quality of the law-systems developed for those ends.

A similar analysis can be performed for proposals aimed at combatting fake news on social media. The proposed Online Safety Bill incorporates many of the same structures deployed in existing voluntary arrangements like the E.U. Code of Practice and presages some of the conditional liability proposals emerging in the U.S., so a closer look at the Online Safety Bill in particular can shed light on other pending regulatory developments as well. Following Reed’s approach, law-system quality can be evaluated by the ability of the law’s subjects to comply with imposed requirements and the effectiveness of the law at achieving its intended results.¹⁰²⁸ With regard to the Online Safety Bill, the relevant considerations are as follows: what compliance with the law looks like, and what results the law aims to achieve.

Regarding the first consideration, compliance with the responsibilities set forth in the Bill includes a wide range of behaviours that social media companies would have to enact. Among these are completing risk assessments and incident reports with regard to regulated content, filing annual transparency reports, offering users digital identity verification services, instituting ‘proportionate’ content moderation and treatment systems, promoting media literacy, and sharing data and consulting with Ofcom’s advisory committee on misinformation.¹⁰²⁹ Some of these compliance aspects reflect the ‘tick-box’ characteristic Reed identified, in particular, the reporting and consulting requirements, the digital ID provision, and media literacy interventions. Other aspects like instituting ‘proportionate’ systems seem to provide the sort of objective measurement social media companies could

¹⁰²⁵ Reed (n 1024) 904-05

¹⁰²⁶ *ibid*

¹⁰²⁷ *ibid* 905-10

¹⁰²⁸ *ibid* 904

¹⁰²⁹ Comparable provisions of the Digital Services Act Regulations include Article 15 ‘Transparency reporting obligations for providers of intermediary services’, Article 20 ‘Internal complaint-handling system’, Article 34 ‘Risk assessment’ and Article 35 ‘Mitigation of risks’, as well as Article 40 ‘Data access and scrutiny’, Article 41 ‘Compliance function’ and Article 42 ‘Transparency reporting obligations’ for very large online platforms

use to reliably comply with the regulation.¹⁰³⁰ However, as Reed noted in the case of the e-signature law, the legal certainty associated with such a provision may conceal an uncertain process. Lawyers would, for example, have to interpret ‘content of democratic importance’ to the technical experts who will have to adjust moderation systems, and the lawyers will then have to evaluate whether the technician’s implementation will be viewed as proportionate. Ofcom would then have to judge the system’s compliance based on the reports they receive from the platforms.

Such questions touch on the second consideration of law-system quality, what result the law aims to achieve, because they throw into relief the fact that the aims of the Online Safety Bill are somewhat diffuse and difficult to pin down. The Bill does not say which treatments the social media companies should apply to what content, for example: their attorneys will have to judge whether the company can comply by merely down-ranking a controversial post about political news—or whether to err on the side of caution and tell the company’s technicians to take the post down altogether.¹⁰³¹ Likewise, while ‘safety’ is in the name of the Bill, what the Bill aims to keep social media users ‘safe’ from is not necessarily apparent. Illegal content like child pornography is already illegal, as is offline violence and tort malfeasance. Other provisions like the digital ID system seem to bear only on identifying social media users—no doubt helpful for certain government aims like surveillance and law enforcement, but which could imperil those whose safety benefits from online anonymity, like children and political dissidents.¹⁰³²

Consultation between Ofcom, the companies and their users, and misinformation experts, another provision of the Online Safety Bill, certainly could help create the kind of interpretive conversation that promotes effective regulation.¹⁰³³ However, adequately incorporating the perspectives of the millions of social media users who produce and consume news content would be a very different prospect from Ofcom sitting down for a

¹⁰³⁰ The Digital Services Act Regulations Article 14 ‘Terms and conditions’ sets out similar expectations for platforms in the E.U. market: policies and procedures must be set forth in the platforms’ terms of service and applied and enforced “in a diligent, objective and proportionate manner,” giving “due regard to the rights and legitimate interests of all parties involved, including the fundamental rights of the recipients of the service, such as the freedom of expression, freedom and pluralism of the media, and other fundamental rights and freedoms as enshrined in the Charter.”

¹⁰³¹ While the Digital Services Act Regulations Article 16 ‘Notice and action mechanism’ sets out precise guidelines for user initiated *reporting* of potentially illegal content, potential responses by the platforms contemplated in Article 17 ‘Statement of reasons’ include the entire gamut of visibility restriction, demonetization, suspension, and termination.

¹⁰³² The Digital Services Act Regulations has its own particularly government-interested provision in Article 36 ‘Crisis response mechanism’ which contemplates the prevention, elimination, or limitation of any “contribution” to a “serious threat to public security or public health in the Union or in significant parts of it”—under “extraordinary circumstances” not to exceed three months (unless extended by another three months by the Commission).

¹⁰³³ Reed (n 1024) 924-26

chat with the heads of the BBC, ITV, and Channel 5.¹⁰³⁴ Broad portions of the user base, such as those who turn to social media news for alternative views, may simply disagree with the legislative designations or with the advisory committee. The Bill does not appear to allow Ofcom or the social media companies to accede to the wishes of the user base in such a case. If users see the consultation process as merely perfunctory, they likely will not view certain enforcement decisions as legitimate.¹⁰³⁵ As in the case of reporting technical compliance, the platforms will have a substantial advantage in representing their own perspectives to the regulator.

Based on these considerations, the Online Safety Bill may be difficult to comply with apart from its *pro forma* aspects and uncertain as to what aim it advances with respect to online content. Like other past cyberspace laws examined by Reed, it exhibits some of the characteristics of a low-quality approach. Regulatory proposals like the Online Safety Bill may ultimately worsen the prospects for improved public trust in online media due to their complexity and lack of clear aims. Where aims and obligations are unclear, the platforms may over-correct by intensifying moderation, which in turn can tend to draw in genuine expression as suggested by my empirical research.¹⁰³⁶

More generally, regulation obligating additional content moderation may potentially worsen the status quo situation with respect to free expression. A duty of care like that under the proposed Online Safety Bill would require social media companies to institute additional systems to address specified content categories like ‘journalistic content’ and ‘content of democratic importance’ in addition to ‘harmful’ content.¹⁰³⁷ These expanded systems necessarily mean more content moderation, and the scale of the task necessarily means that the social media companies will continue to rely on automation. More content moderation in general increases the likelihood that genuine content will be subject to moderation—which in turn increases the likelihood that existing tendencies towards false positives will be reproduced along the lines described above. In order to meet their regulatory obligations, social media companies will necessarily need more data about the content on their platforms. One expedient and technically feasible way for the platforms to obtain that data will be to offer users tools (such as a button in the graphical user interface to click) to flag content according to the categories suggested by the relevant regulation: as

¹⁰³⁴ The Digital Services Act Regulations Article 61 ‘European Board for Digital Services’ and Article 62 ‘Structure of the Board’ establish an “independent advisory group” (composed of the Digital Services Coordinators and chaired by the Commission) to advise the Digital Services Coordinators and the Commission; the Board “may” cooperate with external experts or consult interested parties but is not obligated to include representatives of the platforms or the users.

¹⁰³⁵ Reed (n 1024) 924-25

¹⁰³⁶ Part IV, ‘Findings’ and ‘Implications’; see also Helm and Nasu (n 829) 309-10 (discussing how Germany’s *NetzDG* law has led to overcorrection by the platforms)

¹⁰³⁷ Discussed in ‘Public Policy and the Public Sphere’ – European Perspectives

journalistic, as democratically important, as harmful but not illegal, or as misinformation (which Twitter now offers in select markets).¹⁰³⁸ My empirical research suggests that in the case of political news stories, inputs from users about that content may inevitably reflect a degree of psychologically motivated appraisal.

The heightened legal consequences for failing to meet a duty of care would further pressure social media companies to take steps that increase the likelihood of false positives. One step that social media companies may take to comply with a regulatory duty of care would be to lower the thresholds upon which they act with regard to subject content. The platforms currently use factors like crowd-sharing behaviour or rates and volume of user flagging to screen content for potential moderation. Accordingly, platforms may act in response to pressure to avoid false negatives—the regulatory non-compliance event—by reducing the amount of signal required from these factors for content to enter the automated moderation process. In other words, fewer flags in a shorter timeframe might trigger moderation for a user’s post than before regulation. But because these thresholds are agnostic to the actual veracity of a particular post, genuine news content will also be pulled into moderation and face increased risk of wrongful suppression as discussed in Part Four.¹⁰³⁹

Another step the platforms might institute to avoid breaching their regulatory duty would be to increase the statistical ‘weight’ they assign to comparison factors in their content models. As discussed in Part Two, some of the routinely used factors like lexical features or source ratings contain their own inherent limitations.¹⁰⁴⁰ As fake news and genuine news become increasingly stylistically similar over time, automated detection systems built around stylistic distinction become less accurate in classifying content as fake or genuine. Increasing the weight of shared lexical features would only draw in more of both kinds of content. As more content is drawn in, the tendency of social media platforms may be to err on the side of caution—that is, to avoid regulatory penalty—by treating a greater amount of genuine content as suspect.

Regulatory approaches can also tend to advantage specific private interests, such as the established news media. The Online Safety Bill’s ‘recognized news publishers’ and SB 7072’s ‘journalistic enterprises’, as with the policy determinations of the social media companies concerning fact-checkers and ‘reliable’ sources, would tend to favour existing

¹⁰³⁸ Twitter, ‘How we address misinformation on Twitter’ (2023)

<<https://help.twitter.com/en/resources/addressing-misleading-info>> accessed 04 January 2023

¹⁰³⁹ Part IV, ‘Implications’

¹⁰⁴⁰ Part II, ‘Platform Policies and Technical Capabilities’ – Evaluating Content Moderation at Scale

mass media outlets over independent and amateur news sources.¹⁰⁴¹ As discussed in Part Two, the effect of platform policies which incorporate third-party fact-checkers like NewsGuard or the Atlantic Council, for example, has been the entrenchment of materially interested discourses at the expense of domestic and foreign competitor narratives.¹⁰⁴² Specific regulatory measures prioritizing or shielding content from mass media outlets risk undermining the aspects of social media that make it so useful to public: access to alternative views and news from a diverse range of sources from around the world.¹⁰⁴³

Accordingly, emphasizing a factor like source ratings would tend to push novel and independent sources of news content in to moderation.¹⁰⁴⁴ Signals about sources likewise tend to degrade and shift over time, especially during the turbulent political news cycles surrounding elections. By privileging established or licensed journalistic enterprises, regulation like the Online Safety Bill or Florida's SB 7072¹⁰⁴⁵ would codify the signal upon which social media platforms could rely in distinguishing between trustworthy and untrustworthy sources. This regulatory thumb on the scale would compound existing tendencies in machine learning models to draw independent media and individual user content into moderation while passing on mass media content—again, without regard to the actual veracity of the content any of these players are producing.

Regulatory and co-regulatory approaches may likewise introduce further incentives for the platforms to reach certain determinations about content that benefit the interests of the administrative state over those of the broader societal whole. In the previous section, I described how social media platforms like Twitter and Facebook exhibit vulnerability to government pressure, especially from security agencies.¹⁰⁴⁶ The threat of noncompliance penalties from a regulator may add to the pressure for social media platforms to align their content determinations with state interests to avoid potential sanction. And as discussed in the preceding section of this Part, licensure and censorship bureaucracies served as powerful

¹⁰⁴¹ The Digital Services Act Regulations Article 22 'Trusted flaggers' similarly privileges existing sources of "expertise," although the Digital Services Coordinator makes the determination of who is trustworthy rather than the platforms.

¹⁰⁴² Part II, 'Platform Policies and Technical Capabilities'

¹⁰⁴³ Mitchell observed that the impetus for radio content regulation could have been substantially obviated by actual diversity of station ownership (essentially what a platform like YouTube *could* offer content producers): "The fact that there are potentially 7,000 different radio station owners is a rather remarkable potential. If you actually *had* 7,000 different people owning radio stations, the diversity in approaches to news and opinions would be remarkable . . . we would probably relax the Fairness Doctrine and similar nitty-gritty types of regulation designed to control content. Diversity of ownership would insure considerable diversity of content and approaches to views." Lee M Mitchell, quoted in Dan Lacy, 'Rules for the Regulators: A Rationale for Government's Role' in William L Rivers & Michael J Nyhan (eds), *Aspen Notebook on Government and the Media* (Praeger London 1973) 93, 103

¹⁰⁴⁴ Counter to the consumer preferences discussed in Part I, 'Social Media News Use'

¹⁰⁴⁵ As discussed in the preceding section, both of these approaches create special circumstances around certain news outlets, albeit through rather different definitions of the relevant enterprise characteristics.

¹⁰⁴⁶ See also Part II, 'Platform Policies and Technical Capabilities'

resources for authorities in the pre-regulatory era. They provided early nation states with a legal tool for disciplining discourses that ran counter to the prevailing wisdom of the sovereign. The Tudors, however, did not need the window-dressing of ‘battling impediments to the “missions assigned to the [Homeland Security] agency by law”¹⁰⁴⁷ to justify excluding Catholic perspectives from public life. Where modern governments derive authority from public trust and expertise, heterodox accounts of political news events (transmitted over social media) carry the same systemic threat where they undermine the government’s perceived competence. Provisions of the Online Safety Bill like the digital ID system or other online de-anonymization proposals¹⁰⁴⁸ would necessarily provide the state with further avenues for rooting out dissent.

Traditional Alternative: Judicial System Control

Other proposals aim to empower the judiciary to determine responsibility for social media news content, essentially reasserting the traditional role of the courts in deciding questions of legally permissible speech. Such proposals encompass a range of alternatives, including criminal or civil liability for certain social media communications, but generally rely on the judicial process to identify offending content and hold its communicator to account for its resulting harm. The judicial approach operates under the conventional premise of tort or criminal law—the threat of penalty is expected to deter the creation or dissemination of fake news.¹⁰⁴⁹ The availability of existing common law torts like defamation and fraudulent misrepresentation means that courts do not need to wait for legislatures to pass a particular statute on fake news: judges and parties instead need triable cases and reasoned determinations of what appropriate relief looks like (likely equitable or declaratory rather than damages alone).¹⁰⁵⁰ The approach has the advantage of getting ahead of potentially harmful fake news through its deterrence effect,¹⁰⁵¹ rather than relying on post hoc measures to curb the content (as is currently done by the platforms and as would be further required under proposed regulation like the Online Safety Bill). The time-lag in

¹⁰⁴⁷ Part V, ‘Public Policy and the Public Sphere’ – Trans-Atlantic Perspectives’

¹⁰⁴⁸ McSherry, York, and Cohn (n 1022) (discussing the discursive role of anonymity and fallout of ‘real name’ policies)

¹⁰⁴⁹ Helm and Nasu note however that the efficacy of this premise in the social media context has yet to be demonstrated by empirical study; Helm and Nasu (n 829) 323

¹⁰⁵⁰ Cf. Antonin Scalia, quoted in Lee M Mitchell, ‘Who Watches the Watchdog?: The Crisis of Public Confidence’ in William L Rivers & Michael J Nyhan (eds), *Aspen Notebook on Government and the Media* (Praeger London 1973) 144, 161-62 (then-future Justice Scalia expressed skepticism about how such adjudications would be publicly received, citing the traditional avoidance by courts of politically charged questions of news bias or unfair coverage: “Here you’re getting into an area where our society has made a determination that you can’t adjudicate the rights and wrongs. Surely the reason that the Constitution would prevent the government from suppressing an unfair paper is not because unfair newspapers are good but because it’s too hard to decide what’s fair and unfair. We have no confidence in our ability to decide that question.”)

¹⁰⁵¹ Helm and Nasu (n 829) 323

judicial responsibility instead arises from the challenges of having a justiciable case heard in a timely fashion—particularly when a judicial determination is sought in advance of a pending election. But the clearest advantage of the judicial approach lies in the way it establishes accountability through a public, transparent, and adversarial process.

As discussed previously, the nature of the processes relied upon by the social media companies create substantial challenges for the platforms in articulating how they identify content as fake news and how they determine it to be harmful. Further, social media users must appeal content decisions by the platforms *to the platforms*, which as discussed previously have a host of incentives to decide in their own interests. Courts, on the other hand, rely on a process that has specifically evolved to meet normative and constitutional requirements that embody societal values.¹⁰⁵² Accordingly, any application of the judicial process must be done in a tailored way: the law must make clear what is prohibited.¹⁰⁵³ A court would then assign responsibility pursuant to a specific offence or cause of action, allowing for a contested factual examination, and hearing defences that safeguard against abusive assertions of falsity.¹⁰⁵⁴ Courts thus provide independent structures for determining accountability that aren't tied to the business practices or relationships intrinsic to the social media platforms, and these structures could stand as correctives alongside automated moderation or as replacements in specific roles like fake news management.

Courts also have institutional knowledge in handling matters that are functionally similar to misleading political news, such as cases of fraud or misrepresentation, defamation of public figures, and incitement to political violence. This knowledge has provided flexibility and adaptability even as communications technologies have changed over time. Lavi points out that technology can outpace both the legislative process and regulation, whether through innovation or obsolescence.¹⁰⁵⁵ Speech torts have proven resilient, however, thanks to their sensitivity to social context—whether a statement is defamatory can be considered in relation to the strength of ties and speaker, subject, and hearers, for instance.¹⁰⁵⁶ The potential responsiveness of courts, however, is tied to factors like the availability of room on crowded dockets, triable cases that can advance legal rules, and other aspects of access to justice that complicate the efficacy of the judicial system more broadly. McSherry, York, and Cohn point out that existing election advertising and foreign agent registration laws could be

¹⁰⁵² See Helm and Nasu (n 829) 323-24

¹⁰⁵³ *ibid* 309-11, 323-24

¹⁰⁵⁴ *ibid* 325

¹⁰⁵⁵ Michal Lavi, 'Content Providers' Secondary Liability: A Social Network Perspective' (2016) 26(4) Fordham Intellectual Property, Media & Entertainment Law Journal 855, 981

¹⁰⁵⁶ *ibid* 892

better brought to bear in dealing with the most pressing concerns about fake news interference.¹⁰⁵⁷

Applying the judicial process to social media interactions would also require the resolution of a substantial number of uncertainties. First and foremost would be clarifying the role of the platforms in judicial proceedings, given that the companies currently enjoy their unique liability protections under provisions like Section 230. Newspapers and other media industries have long histories of involvement in tort and criminal cases, whether as parties or in providing the kinds of evidence germane to such proceedings. Social media companies would have to develop legal competencies of their own, and judges and attorneys would have to figure out what sorts of evidence about social media use bears on a case involving fake or misleading news content. Lavi suggests that the interactive context of social media can be readily categorized in sociological terms that could be used in formulating liability rules and describing the scope of harmful online speech.¹⁰⁵⁸ But to describe those interactions accurately, lawyers will need to seek some of the same information from the platforms that I argued in Part Four will likely remain opaque to individual users.

To that extent, judicial proceedings would rely on the platforms' responsiveness to subpoenas and other evidence gathering measures. The social media platforms may be specifically advantaged over other parties in some of these regards, especially owing to their substantial financial resources. Plaintiffs and defendants alike could face significant obstacles to producing adequate testimony about how other users perceived the content in question: potential witnesses—the other users—may be scattered across the planet rather than close to the judicial venue, and their credibility accordingly may be difficult to assess without reference to the kinds of community ties and personal knowledge that traditionally guide finders of fact. Nevertheless, networks like Facebook may feature enough routine users with strong social ties (and repeated interactions) with one another¹⁰⁵⁹ to support established evidentiary processes.

The risks to free speech and the press inherent in a judicial approach to deciding responsibility for social media content arise primarily from vague or overly broad laws and arbitrary enforcement.¹⁰⁶⁰ Maintaining a specific focus on social media news in the elections context may promote the development of sufficiently tailored laws. By limiting the scope of government interference to specific time periods around elections and requiring specific

¹⁰⁵⁷ McSherry, York, and Cohn (n 1022)

¹⁰⁵⁸ Lavi (n 1055) 894-95

¹⁰⁵⁹ *ibid* 905-06

¹⁰⁶⁰ Helm and Nasu (n 829) 323-24

showings of intent and likelihood to disrupt election proceedings, legislatures could potentially provide courts with proportionate laws that squarely balance potential harms and government interference with social media news use.¹⁰⁶¹ The time pressure of elections further complicates the application of existing judicial resources like temporary restraining orders and injunctions: in order to quickly deal with fake news bearing on an election, such tools are an obvious bridge between present harms and future deliberations. But how courts determine their issuance varies across jurisdictions and the equities at stake will necessarily involve political elections. Furthermore, while courts have the institutional independence to evaluate laws and make determinations about their constitutionality, their exercise of this prerogative is not necessarily neutral, nor guaranteed. Judges will interpret laws about speech over social media in keeping with their own understanding of legitimate government aims and interests, just as they have in the past.

Beginning in 2019, the U.S. Department of Justice (‘DOJ’) undertook a review of Section 230, concluding that its provisions may no longer be well fitted to the ways the Internet is used more than twenty-five years after the law was adopted.¹⁰⁶² The DOJ, in its consultations with experts, industry, lawmakers, and the public, identified the following reform proposals: adjusting the incentives platforms have to address illicit content, clarifying civil enforcement capabilities, promoting competition through antitrust measures, and encouraging transparency between the platforms and users.¹⁰⁶³ Conditional liability protection schemes appear to be more in line with the DOJ’s recommendations, which continue to stress the contribution of liability protection to a “vibrant, open, and competitive” Internet,¹⁰⁶⁴ rather than a complete repeal of the Section 230 provisions.

The DOJ reform approach focuses on facilitating the Department’s law enforcement responsibilities while attempting to avoid undue burden on social media platforms.¹⁰⁶⁵ As such, the DOJ does not propose to make online platforms the arbiters of true or false statements in instances of alleged defamation.¹⁰⁶⁶ The DOJ instead seeks carve-outs from liability protection for “bad actors” that promote or facilitate illegal conduct online, such as terroristic activity or child sexual exploitation and cyberstalking.¹⁰⁶⁷ In its carve-out proposals, the DOJ recognizes that absent the blanket liability protection of Section 230, platforms would have to meet traditional tort standards with respect to terrorist content or

¹⁰⁶¹ Helm and Nasu (n 829) 324

¹⁰⁶² U.S. Department of Justice, ‘Section 230 – Nurturing Innovation or Fostering Unaccountability?’ (June 2020) <<https://www.justice.gov/file/1286331/download>> accessed 28 July 2022 (hereinafter ‘DOJ Report’)

¹⁰⁶³ *ibid* 3-4

¹⁰⁶⁴ *ibid* 5; see, *e.g.*, Ending Support for Internet Censorship Act, S. 1914 (116th Congress 1st session, 19 June 2019)

¹⁰⁶⁵ DOJ Report (n 1062) 9

¹⁰⁶⁶ *ibid* 11-12

¹⁰⁶⁷ *ibid* 14-16

child sexual abuse material.¹⁰⁶⁸ Accordingly, the DOJ proposes that notice liability (actual knowledge in cases of specific criminal material) could incentivize platforms to address such material promptly without placing the platforms in the same position as publishers, who would be strictly liable for defamation they publish.¹⁰⁶⁹ Platforms already comply with notice-and-takedown requirements in copyright law, which is carved out of Section 230.¹⁰⁷⁰ The DOJ argues that notice liability avoids concern of platforms being forced to respond to any “heckler’s veto” on users’ posted content by putting the adjudication process in the hands of the courts.¹⁰⁷¹

Although the DOJ report does not explicitly discuss elections-related speech content or fake news, the DOJ approach turns on the legislative determination to criminalize certain speech and the judicial review of that determination as constitutional or not, which some proposals related to fake news content on social media seek to do. However, the DOJ does specify its support for limiting the discretion of platforms to moderate content that is merely “otherwise objectionable” as provided under Section 230 currently.¹⁰⁷² The report notes that “unconstrained discretion is particularly concerning in the hands of the biggest platforms, which today effectively own and operate digital public squares.”¹⁰⁷³ In this regard, the DOJ approach urges greater transparency from the platforms in their ‘terms of service’ provisions—the rules by which the platforms require users to follow in using their services.¹⁰⁷⁴ The terms of service provisions would continue to play an important role in platforms’ liability protection for their moderating decisions, so the DOJ urges legal clarification for what constitutes “good faith” abidance by the platforms with those provisions as well as public disclosure requirements on content moderation policy enforcement.¹⁰⁷⁵

How then might we expect judges to handle a statute that would, for example, criminalize election-period false statements intended to interfere with elections or discourage voting that are transmitted over social media? History shows that the process of delineating clear boundaries for the scope of government interference with speech in other can pass constitutional muster but still result in arbitrary prosecutions. The 1917 Espionage Act¹⁰⁷⁶ criminalized wartime false statements intended to interfere with U.S. military

¹⁰⁶⁸ DOJ Report (n 1062) 16

¹⁰⁶⁹ *ibid* 17

¹⁰⁷⁰ 17 USC § 512(c)(3); 47 USC § 230(e)(2)

¹⁰⁷¹ DOJ Report (n 1062) 17-18; see also Tarleton Gillespie, ‘Governance of and by platforms’ in Jean Burgess, Alice Marwick, and Thomas Poell (eds), *The SAGE Handbook of Social Media* (SAGE 2018)

¹⁰⁷² DOJ Report (n 1062) 21

¹⁰⁷³ *ibid*

¹⁰⁷⁴ *ibid* 21-22 (‘terms of service’ provisions are discussed further in a subsequent section of this Part)

¹⁰⁷⁵ *ibid* 22, 25

¹⁰⁷⁶ 65 PUB. L. 24 [1917]

operations, attempts to incite disloyalty among the armed forces, and the obstruction of military recruiting and enlistment.¹⁰⁷⁷ The penalty for the offense included a fine of not more than \$10,000 or imprisonment for not more than twenty years, or both—a remarkably punitive fine considering that the vast majority of annual incomes in the US in 1917 amounted to less than \$1,000.¹⁰⁷⁸ The Espionage Act also banned the use of the U.S. Mail to send any letter, newspaper, pamphlet, book or other media that violated any other provisions of the Act as well as any media urging treason, insurrection, or forcible resistance to any law of the United States.¹⁰⁷⁹ These provisions sparked a series of important free speech cases heard by the U.S. Supreme Court that illustrate the complexity of developing effective judicial rules concerning context, intent, and the legitimate use of government power.¹⁰⁸⁰

As with the Sedition Act and Espionage Act, new communications offences could increase the risk of arbitrary prosecutions. In the proposed Online Safety Bill, arbitrary application is written into the statute: the Section 160 false communications offence specifically exempts its application to ‘recognized news publishers’ and licensed broadcasters.¹⁰⁸¹ Individual users, who probably attract relatively smaller audiences and possess relatively fewer resources to withstand prosecution than mass media outlets, are apparently exclusively responsible for the consequences of false communications and will bear the full brunt of government intervention. The political nature of elections news only worsens the temptation of authorities who may prefer to suppress opposing speech rather than let their perspectives withstand public debate.

Allocating responsibility to the judiciary for determining accountability for social media news content can provide distinct advantages in process transparency, independence, and adaptability compared to reliance on other institutions. But realizing these advantages would depend on addressing the factors that could diffuse or misdirect its time-tested processes. As new laws attempt to create offences or causes of action resulting from online speech, judges will have to make historically familiar determinations as to what social media

¹⁰⁷⁷ 65 PUB. L. 24, s. 3; 65 PUB. L. 150 [1918] (the 1918 amendments added to the list of Section 3 offences: utterance or publication of any “disloyal” abuse of the government, statements intended to bring the government or military into contempt or disrepute, incitement or encouragement of resistance to U.S. authorities, discouragement of the production of war materiel, and the advocacy, teaching, or defence of any of the Act’s enumerated offences)

¹⁰⁷⁸ 65 PUB. L. 24, s. 3; United States Internal Revenue, *Statistics of Income* (GPO, 1919) 24
<<https://www.irs.gov/pub/irs-soi/17soirepar.pdf>> accessed 6 March 2022

¹⁰⁷⁹ 65 PUB. L. 24, Title XII; 65 PUB. L. 150 (To Title XII of the Espionage Act, the 1918 amendment added a further provision granting the Postmaster General sweeping power to interdict and return to the sender any mail addressed to a person or organization, “upon evidence satisfactory to [the Postmaster],” suspected of using the postal service in violation of any of the Act’s provisions)

¹⁰⁸⁰ Additional detail on these cases is provided in **Appendix D**

¹⁰⁸¹ §§ 160, 161 Online Safety Bill, Bill 220 2022-23; Florida’s SB 7072 contains similar exceptions

contexts give consequence to misleading or fake political narratives. They will have to decide whether social media users posted such stories with the requisite intent and whether a resulting harm is so serious as to justify the government infringing the rights of its citizens. And they will depend on governments and plaintiffs to promptly and judiciously bring cases with actual merit in courtrooms with the resources and docket space to manage such an undertaking.

Radical Traditional Alternative: Legislative Control

A further alternative to reliance on automated moderation by large social media platforms could draw on historical examples to radically reconceptualise the social media platforms as *infrastructure*, focusing on open access along constitutional principles. Allocating authority over social media content to a legislative body like Parliament or Congress requires something different than making laws that empower a regulator to make further rules for the platforms or the courts to make accountability determinations based on a criminal statute. Instead, legislatures might take more direct responsibility for social media content by creating or reshaping structures (particularly the relationship between platforms, users, and news) to function a certain way. While such structures may rely on corporations, the administrative state, or judiciary to ensure their operation, legislative bodies are unique in their power to directly express and reformulate normative aims on behalf of the public by establishing or abolishing institutions. Such power was evident in the Congress's initial action to create a new source of private power under Section 230 for making decisions about user content, rather than allocating that authority to courts as with newspapers or a regulator as with broadcast media. An approach that establishes accountability for content through legislative power would thus require the legislative body to articulate a clear normative aim for the social media system and, if necessary, create or reform the structures of that system in line with the intended function.

Such an approach could take any number of forms, but legislatures (like courts) have an extensive background to draw on in considering what might work to achieve particular goals like the promotion of public discussion of the news. This background includes public support for news distribution, which could inform approaches that aim to create publicly owned social media platforms as alternatives to the existing private corporation platforms, as well as prohibition of access discrimination, which could inform approaches that aim to require common carriage of user content over social media channels. Recent attempts at legislative reforms of content moderation practices in Florida and Texas provide further

insights into some of the key constitutional questions that legislative bodies must answer with a clear vision of how social media operates in relation to broader society.¹⁰⁸²

Public subsidy of distribution channels—how the citizenry gets its news—has long been one of the most important ways in which democratic liberty has traditionally been promoted and safeguarded. Public policy decisions in support of libraries, for example, have been touted as important contributions to disseminating ideas, as have laws like the 1879 Postal Act that “provided for the free transportation in-county for newspapers and other publications as well as a highly preferential national postal rate for magazines.”¹⁰⁸³ Lacy, a vice president of the publisher McGraw-Hill, observed the following:

“This kind of affirmative government action that affects the distribution machinery becomes more and more important. You can’t look at the First Amendment only in its absence of restraint. You have to think about the structure. It may involve as diverse things as copyright in certain cases, as library support, as postal rates, as access in broadcast media, as rural free delivery. It doesn’t sound very relevant, but it was a tremendously important thing in achieving truly national communication.”¹⁰⁸⁴

After the Revolutionary War, the legislative power to establish post offices and roads was made part of the U.S. Constitution,¹⁰⁸⁵ and the first major postal law was passed by the Congress in 1792 to promote a free press by allowing newspapers “to travel through the U.S. Mail at extremely low rates of postage – in some cases for free – to ensure the success of the democracy.”¹⁰⁸⁶ As noted by Kielbowicz, Postmasters General periodically complained to the Congress about the expense of carrying news between editors without charge; however, the Congress preserved the exchange privilege well into the telegraph era.¹⁰⁸⁷

Recognizing that the social media platforms exhibit similar infrastructural dynamics opens up similar policy options that have successfully promoted infrastructure like postal services in the past.¹⁰⁸⁸ Akin to the public subsidy for news distribution provided through

¹⁰⁸² Discussed further below

¹⁰⁸³ Dan Lacy, quoted in Roger Fisher, ‘Press Rights vs. Press Responsibilities: The Issues of the First Amendment’ in William L. Rivers & Michael J. Nyhan (eds), *Aspen Notebook on Government and the Media* (Praeger London 1973) 51, 67

¹⁰⁸⁴ *ibid*

¹⁰⁸⁵ U.S. CONST., art. 1, s. 8

¹⁰⁸⁶ U.S. Postal Service, *The United States Postal Service: An American History* (Washington, D.C. 2020) 1

¹⁰⁸⁷ Richard B Kielbowicz, ‘News Gathering by Mail in the Age of the Telegraph: Adapting to a New Technology’ (January 1987) 28(1) *Technology and Culture* 26, 30

¹⁰⁸⁸ Rahman (n 1006) 1641-45 (identifying factors like returns on scale, downstream activity the service enables, social necessity, and bottlenecks like private control over terms of access that characterize infrastructure)

the operation of a public postal service, legislators could create a ‘public option’ for social media. A state run platform would have to abide by constitutional principles in its treatment of users and content, and provide adequate processes to reflect those principles. Additionally, a public option would not need to be subject to advertising incentives; a public social media platform could be funded by taxes or subscription, thereby avoiding one driver of fake news content¹⁰⁸⁹ from its inception. Users would become the customers instead of pawns in a two-sided market, and would have recourse for customer dissatisfaction through their elected representatives who oversee the service—if nothing else, a marginal improvement over the current relationship. A public option would compete with existing platforms, potentially encouraging innovation and process improvements on the part of private firms, but would also face barriers to adoption like switching costs and network effects that tend to keep users from leaving existing platforms.¹⁰⁹⁰ Nationalisation of existing social media platforms would represent the extreme end of providing a public option, obviating any concerns about competition or network effects, but would also entail a substantial deviation from current political orthodoxy.

Nationalisation of a communications system is not without precedent, although the chief example benefitted from its accomplishment relatively early in the infrastructural development of the technology. Britain took decisive action on the telegraph much earlier than did the United States: As Victorian reformers pushed for legal changes promoting market competition in news, a change that helped achieve this aim was, perhaps counter-intuitively, to nationalize telegraphy in 1868–1870.¹⁰⁹¹ By making telegraphy service equally available to all users and more readily affordable, the British government apportioned distribution and circulation costs in such a way that local papers in outlying areas could benefit from most, similar to the way other nationalized network industries like railways worked.¹⁰⁹² Competition among city and provincial publishers increased in the absence of telegraphy cartels,¹⁰⁹³ and the British press in general was better positioned to bargain with outside monopolists like Reuters that provided foreign news.¹⁰⁹⁴ But as noted by Silverstein-Loeb, such an arrangement was not without trade-offs: pricing structures were often complex and the Post Office bore the administrative and technical burden of operating the

¹⁰⁸⁹ Adam J Mills, Christine Pitt, and Sarah Lord Ferguson, ‘The Relationship between Fake News and Advertising: Brand Management in the Era of Programmatic Advertising and Prolific Falsehood’ (March 2019) *Journal of Advertising Research* 3, 4-5

¹⁰⁹⁰ Doctorow (n 1005); see also Rahman (n 1006) 1679 (noting the importance of public utilities not merely recreating existing inefficiencies)

¹⁰⁹¹ Jonathan Silberstein-Loeb, ‘The Structure of the News Market in Britain, 1870–1914’ (Winter 2009) 83 *Business History Review* 759, 760-61

¹⁰⁹² *ibid*

¹⁰⁹³ *ibid* 763-64

¹⁰⁹⁴ *ibid* 770-72

telegraphs equally and adequately as *de facto* common carrier.¹⁰⁹⁵ Yet in facilitating the wider distribution of news, the nationalization policy successfully achieved its aims.¹⁰⁹⁶

Understanding social media platforms as infrastructure further highlights the relevant public values at stake, like non-discrimination and broad access to information.¹⁰⁹⁷ An absence of that understanding characterized American public policy during the latter half of the 19th Century, along with a certain mysticism about new technology. While subsidy through the post office had been a boon to the local and upstart newspapers of the U.S. in earlier decades, the mid-19th Century shift favouring an unregulated communications network through the telegraph would prove to be decidedly less effective in sustaining a vigorous American free press.¹⁰⁹⁸ Court cases sought to establish the legal responsibilities to the public of the telegraph companies not to discriminate in their carriage of messages, with California in 1859 holding them to be common carriers.¹⁰⁹⁹ But in most other early cases, the courts,

“when they considered the nature and power of electricity, thought it so strange, wonderful and incomprehensible that no ordinary human care or skill could possibly suffice to control it perfectly, and, deeming it therefore unjust to hold telegraph companies bound by the strict rules which govern common carriers, sought out reasons for making a distinction between these new carriers of thought and the old carriers of merchandise.”¹¹⁰⁰

Dozens of bills to reform the telegraph system were considered in the later decades of the nineteenth century, including proposals to set up a public competitor operated by the Post Office.¹¹⁰¹ In 1910, after the mysticism surrounding electricity had worn off, the Congress amended the Interstate Commerce Act to place telegraph companies under the regulatory authority of the federal Interstate Commerce Commission as common carriers.¹¹⁰² As Dixon

¹⁰⁹⁵ Silberstein-Loeb (n 1091) 767-68, 772-774

¹⁰⁹⁶ *ibid* 787-88

¹⁰⁹⁷ Rahman (n 1006) 1668-69; see also Anne Branscomb, ‘Citizen Access to the Media’ in William L Rivers & Michael J Nyhan (eds), *Aspen Notebook on Government and the Media* (Praeger London 1973) 120, 120 (providing an extended definition of access to the media that captures relevant relational aspects: “It may mean access to speak, or access to programming that one wishes to view, or access to ownership of the media, or access to the decision-making process. However it is expressed, there is sharp public frustration over the inability of individuals both to control the institutions that dominate their daily existence and to participate in the decisions that directly affect them.”)

¹⁰⁹⁸ Discussed in Part II, ‘Social Media and Media Systems’

¹⁰⁹⁹ Benjamin F Rex, ‘Liability of Telegraph Companies for Fraud, Accident, Delay and Mistakes in the Transmission and Delivery of Messages’ (May 1884) 32(5) *American Law Register* 281, 281-82

¹¹⁰⁰ *ibid* 282

¹¹⁰¹ Czitrom (n 985) 27-28

¹¹⁰² Frank Haigh Dixon, ‘The Mann-Elkins Act, Amending the Act to Regulate Commerce’ (August 1910) 24(4) *Quarterly Journal of Economics* 593, 596

observed at the time, “there seems to be no good reason why the Commission cannot apply to agencies of this character the same rules of action that are applied to transportation companies.”¹¹⁰³

When considered in clear light of the centrality and control exhibited by the social media platforms, common carriage rules present a powerful legislative tool for aligning social media infrastructure with users’ interests in access to diverse news sources.¹¹⁰⁴ Common carriage would require social media platforms to relinquish their power over news circulation to the users. That shift downstream would immediately lessen the influence of the platforms (and by extension, mass media outlets and governments that shape platform content policy) over information that is vital to elections and public discourse.¹¹⁰⁵ This does not mean that the inmates will be running the asylum, however—illegal content like child pornography remains illegal and subject to interdiction and prosecution, just as in other infrastructures like the mail or telephone system, and enforcement of existing laws on elections advertising and foreign agents will be just as important as they are now, as discussed *supra*. In combination with legislative proposals like the ACCESS Act,¹¹⁰⁶ common carriage rules could support the development of new tools for users to manage the content they see. Opening neutral platforms to allow third party network interfaces would preserve the value added by the existing networked user base and infrastructure investments while letting users shop around for the apps they use to connect to the social network. These competing interfaces would empower users to pick what types and level of content moderation they want without having to migrate to a specific platform.¹¹⁰⁷ Unlike the platforms, the app developers would have an incentive to be transparent about their content filtering processes in order to differentiate their offerings.

The court cases resulting from social media laws passed in Florida and Texas highlight the difference between approaches that invade platform practices and those that open up platform structures. The industry group representing the social media companies

¹¹⁰³ Dixon (n 1102) 597

¹¹⁰⁴ Rahman (n 1006) 1670-72 (describing how infrastructural platforms create “a vulnerability among users who could be excluded from access” and a selective presentation of information due to business model and structural factors); see also Ithiel de Sola Pool, quoted in Branscomb (n 1097) 131 (articulating the role of access further: “The principle of the First Amendment is that access should be available to everybody regardless of whether what he says is constructive or useful or wise or influential, or whether it serves any important purpose in the making of national policy.”)

¹¹⁰⁵ Rahman (n 1006) 1670-72

¹¹⁰⁶ ACCESS Act of 2021, H.R. 3849 (117th Congress, 1st Session, 11 June 2021) (requiring covered platforms to maintain interoperable third-party-accessible interfaces)

¹¹⁰⁷ This is fairly in line with the drafting intent of § 230, i.e., letting parents decide what kinds of content they want their children exposed to online—what the § 230 drafters did not anticipate in 1996 was the consolidation of online networking into a handful of giant, closed platforms that forestalled the development of a market for filtering software.

sued to stop the laws in both states. The court challenges presented constitutional questions that turn on whether we accurately view social media as a media system, a situation presaged by Crawford in her analysis of the arguments made by internet service provider Verizon in its 2014 lawsuit against the FCC's 'net neutrality' rules.¹¹⁰⁸ Crawford warned that the power to "selectively present digital communications to users and close their gates to content" would facilitate access providers' arguing that they are more like newspaper editors and less like "a communications transport network subject to economic regulation."¹¹⁰⁹ In response to similar arguments, the federal appellate courts for the respective states recently reached sharply contrasting conclusions about where the balance of regulatory power should lie, as described below.

In the Florida case, the plaintiffs had sought to enjoin the enforcement of SB 7072's content moderation restrictions, including the elections candidate de-platforming provision, and its disclosure obligations.¹¹¹⁰ The federal District Court granted a preliminary injunction, holding that SB 7072's imposition of liability for content moderation decisions would be preempted by Section 230's safe harbour provision.¹¹¹¹ The District Court further held that SB 7072's requirements would impermissibly interfere with the plaintiffs' "constitutionally protected exercise of 'editorial judgment'" without any commensurate governmental interest or sufficient tailoring to avoid unduly burdening the platforms.¹¹¹² The State of Florida appealed, arguing that SB 7072 *as applied* would not necessarily be inconsistent with Section 230, and that the First Amendment challenge was inappropriate because the platforms act alternatively as hosts for the speech of third parties, just as shopping centres or college campuses might do, or else as designated common carriers.¹¹¹³

In its decision, the Eleventh Circuit agreed with NetChoice's contention that the platforms' content moderation decisions are editorial judgments, and that the platforms are more akin to newspapers and parade organizers than private spaces or services open to the public.¹¹¹⁴ The Eleventh Circuit reasoned that if the social media platforms are engaged in First Amendment-protected activity, then the Court must evaluate SB 7072 in that light first and foremost. The Eleventh Circuit decision relied on the Supreme Court's previous evaluations of a 'right of reply' law that compelled newspapers to give political candidates

¹¹⁰⁸ Susan Crawford, 'First Amendment Common Sense' (30 September 2014) 127 Harvard L Rev 2343

¹¹⁰⁹ *ibid* 2345 ("The risk is that a court will someday take the providers' First Amendment position seriously.")

¹¹¹⁰ *NetChoice, LLC v. Att'y Gen. of Fla.*, No. 21-12355 (11th Cir. 2022) ("*Florida*") 13-14; see also SB 7072 § 106.072 and § 501.2041

¹¹¹¹ *Florida* (n 1110) 14

¹¹¹² *ibid* 14-15

¹¹¹³ *ibid* 15

¹¹¹⁴ *ibid* 16 (citing *Miami Herald Publishing Co. v. Tornillo* [1974] 418 U.S. 241 and *Hurley v. Irish-American Gay, Lesbian and Bisexual Group of Boston* [1995] 515 U.S. 557)

equal space to respond to a newspaper's criticism, and of 'must-carry' provisions requiring cable operators to allocate channels to broadcasters.¹¹¹⁵ Determining that these Supreme Court precedents stood for the protection of editorial judgment as an act of expression in itself, the Eleventh Circuit then concluded that some aspects of SB 7072 necessarily interfered with the platforms' preferences to exclude certain participants or particular messages with which the platforms did not agree.¹¹¹⁶ The Eleventh Circuit distinguished the situation of the social media companies from that of shopping malls or law schools (which may be required to host speech with which they disagree) by pointing out where SB 7072 limits the platforms' own speech rights.¹¹¹⁷ The judges point specifically to SB 7072's prohibition against platforms posting an addendum or disclaimer on users' posts¹¹¹⁸ and, owing to an interference with the platforms' editorial judgment, SB 7072's requirement that the platforms host the content of electoral candidates or journalistic enterprises with which the platforms disagree.¹¹¹⁹

The Eleventh Circuit panel then proceeded to evaluate the District Court's determination to treat those provisions as impermissibly viewpoint based—a characteristic which would subject the law to the highest constitutional scrutiny. The Eleventh Circuit did not find support for NetChoice's argument that the entirety of the law was improperly motivated;¹¹²⁰ however, the panel did note the apparent content-specificity of SB 7072's provisions restricting platforms' moderation decisions around the content of posts by journalistic enterprises and those by or about an electoral candidate.¹¹²¹ The Eleventh Circuit determined those provisions faced strict constitutional scrutiny,¹¹²² while the consistent-treatment and other moderation provisions pertained to 'expressive conduct' and only faced intermediate scrutiny.¹¹²³ The disclosure requirements of SB 7072 triggered a more lenient standard.¹¹²⁴ The panel then concluded that the State of Florida had not demonstrated a sufficient governmental interest to sustain its imposition on the platforms' content moderation decisions.¹¹²⁵ As noted in the decision,¹¹²⁶ the appellants apparently did not adequately develop any contentions related to the State's interest in the public's access

¹¹¹⁵ *Florida* (n 1110) 20-23 (discussing *Miami Herald* and the *Turner* cases); cf. Crawford (n 1108) 2364-65 (describing these cases as inapposite to what are functionally conduits for others' speech)

¹¹¹⁶ *Florida* (n 1110) 26

¹¹¹⁷ *ibid* 31-36

¹¹¹⁸ *ibid* 37 (citing SB 7072 § 501.2041(1)(b))

¹¹¹⁹ *ibid* 37-38, 46 (citing SB 7072 §§ 106.072, 501.2041(2)(h), and 501.2041(2)(j))

¹¹²⁰ *ibid* 50-51

¹¹²¹ *ibid* 55

¹¹²² *ibid*

¹¹²³ *ibid* 56

¹¹²⁴ *ibid* 56-57

¹¹²⁵ *ibid* 57-62

¹¹²⁶ *ibid* 58 ("The State seems to have wagered pretty much everything on the argument that S.B. 7072's provisions don't trigger First Amendment scrutiny at all.")

to diverse information generally or in regard to elections specifically.¹¹²⁷ With the exception of the vague ‘thorough’ explanation requirement of SB 7072’s notice provision, the Eleventh Circuit concluded lastly that the Florida law’s disclosure and data access requirements passed constitutional muster.¹¹²⁸

In the Texas case, the same plaintiffs sought a preliminary injunction against HB 20’s prohibition of viewpoint-based censorship as well as its disclosure and operational requirements, including the appeal and review procedures.¹¹²⁹ A federal District Court held that HB 20’s viewpoint provision unconstitutionally interfered with social media platforms’ protected editorial discretion and that the disclosure requirements were unduly burdensome, and accordingly issued a preliminary injunction.¹¹³⁰ Texas appealed and was granted a stay by the Fifth Circuit Court of Appeals, but the stay was subsequently vacated by the U.S. Supreme Court in a 5-4 decision.¹¹³¹ In the report on the decision to vacate the stay, three of the Justices joined a dissenting opinion which noted that while under certain circumstances the Court had recognized the right of organizations to refuse to host the speech of others, as in the case of newspapers, the Court had rejected such claims in other circumstances, as in the case of shopping malls and with cable television operators.¹¹³² The dissent further notes, “It is not at all obvious how our existing precedents, which predate the age of the internet, should apply to large social media companies.”¹¹³³ The appeal proceeded, and the Fifth Circuit reached a subsequent decision that HB 20’s viewpoint-censorship prohibition was not facially unconstitutional and that the plaintiffs were not entitled to pre-enforcement relief against the statute.¹¹³⁴

In its decision, the Fifth Circuit reasoned that a facial challenge to the constitutionality of HB 20 was inappropriate because the relevant provisions of HB 20 do not chill speech but instead pertain to the platforms’ conduct, for which constitutional protections are attenuated.¹¹³⁵ The Fifth Circuit distinguished the conduct of the platforms from those of newspapers, which exercise extensive editorial control over the speech they host,¹¹³⁶ and from parades, where parade organizers’ messages are intimately connected

¹¹²⁷ *Florida* (n 1110) 58-60 (curiously citing as support *Buckley v. Valeo* [1976] 424 U.S. 1, where the Supreme Court recognized campaign contributions as political speech but nevertheless upheld contribution limitations as serving a basic governmental interest in elections integrity)

¹¹²⁸ *ibid* 63-65

¹¹²⁹ *NetChoice, LLC v. Paxton*, No. 21-51178 (5th Cir. 2022) (“*Paxton*”) 5-6

¹¹³⁰ *ibid* 6

¹¹³¹ *ibid* 7

¹¹³² *NetChoice v. Paxton*, 596 U.S. __ (2022) (Alito, J., dissenting)

¹¹³³ *ibid* 3

¹¹³⁴ *Paxton* (n 1129) at 7-8

¹¹³⁵ *ibid* 12-15

¹¹³⁶ *ibid* 21-22 (discussing *Miami Herald*)

with the participants they choose to include.¹¹³⁷ In those instances, the Supreme Court was concerned with compelled speech, and the Fifth Circuit rejected the notion that the platforms' *ex post* censorship of their users' speech is in itself a protected form of expression.¹¹³⁸

The Fifth Circuit found additional support for its conclusions in Section 230, which "reflects Congress's judgment that the Platforms do not operate like traditional publishers and are not 'speaking' when they host user-submitted content,"¹¹³⁹ and in the common carrier doctrine, which permits states to enact laws prohibiting discrimination by certain firms holding themselves out to serve the public.¹¹⁴⁰ Lastly, the Fifth Circuit concluded that even if the platforms' First Amendment rights were implicated by HB 20, the Texas regulations are unrelated to the specific content of the subject speech and therefore faced intermediate constitutional scrutiny.¹¹⁴¹ The Fifth Circuit held that HB 20's viewpoint-censorship prohibition advanced an important governmental interest in protecting the multiplicity of information sources available to the public,¹¹⁴² and that the provision was adequately tailored to the typical content moderation practices affecting users' protected speech on the most relevant social media platforms.¹¹⁴³ The Fifth Circuit likewise rejected the platforms' facial challenge of the reporting and disclosure requirements of HB 20, holding that the requirements were not unduly burdensome to the platforms' protected speech activities.¹¹⁴⁴

The Fifth Circuit reached its decision several months after the publication of the Eleventh Circuit's decision in *NetChoice, LLC v. Att'y Gen. of Fla.* As noted by the Fifth Circuit, Florida's SB 7072 and Texas's HB 20 are substantially different laws, generalizing that "SB 7072 prohibits *all* censorship of *some* speakers, while HB 20 prohibits *some* censorship of *all* speakers."¹¹⁴⁵ But because the panel in *Paxton* reached different conclusions about Supreme Court precedent from those reached in *Florida*, the Fifth Circuit devoted a significant portion of its decision to explaining those differences of judicial opinion. Unlike the Eleventh Circuit, the Fifth Circuit did not find that the Supreme Court had recognized "editorial discretion" as an independent category of First Amendment-protected speech, nor that the social media platforms engaged in the sort of "editorial judgment" the Supreme Court had discussed.¹¹⁴⁶ Further, the Fifth Circuit disagreed with the Eleventh Circuit's determination of whether

¹¹³⁷ *Paxton* (n 1129) 25-26 (discussing *Hurley*)

¹¹³⁸ *ibid* 34-38

¹¹³⁹ *ibid* 39

¹¹⁴⁰ *ibid* 44-45, 53-58

¹¹⁴¹ *ibid* 64-65

¹¹⁴² *ibid* 68-69 (citing *Turner*)

¹¹⁴³ *ibid* 71-72 (further discussing demonetization and de-boosting specifically in fn 35)

¹¹⁴⁴ *ibid* 73-78

¹¹⁴⁵ *ibid* 80-81

¹¹⁴⁶ *ibid* 83-88

common carrier non-discrimination obligations could be imposed on the social media companies.¹¹⁴⁷

Following the publication of the Fifth Circuit’s opinion in *Paxton*, State of Florida officials filed a petition with the U.S. Supreme Court asking for review of the Eleventh Circuit’s judgment in light of the split between the Circuits.¹¹⁴⁸ If granted, the Supreme Court would then look at whether the First Amendment prohibits a State from requiring social media companies to host users’ communications or regulating how they do so, and whether the First Amendment prohibits a State from requiring social media companies to notify and provide an explanation to users about censorship decisions.¹¹⁴⁹ The Supreme Court has also recently granted certiorari regarding a case involving Section 230 and whether the federal law immunizes social media platforms when they make “targeted recommendations” of user-generated content—or only when the companies engage in “traditional editorial functions” regarding such content.¹¹⁵⁰ As these cases proceed, the Supreme Court will provide vital clarity for policy makers attempting to develop constitutionally appropriate rules for news content shared over social media.

Conclusions

Over the course of this Thesis, I have covered key aspects of social media news use in electoral politics and its apparent social dysfunction as described by the ‘fake news’ phenomenon. Proposals for regulatory solutions and the moderation practices implemented by the platforms reflect a broad based concern over how people decide their votes. The approaches taken by public and private policy makers vary in their attention to the psychological drivers and structural factors that influence the creation and propagation of political content online. My empirical research highlights an intrinsic challenge in reform approaches that maintain or rely on existing structures like centralized algorithmic content filtering. Recognizing the limitations of platform-based moderation encourages a closer look at other factors in the social media system that can better be aligned towards improving the information citizens seek in civic participation.

Elections have become the flashpoint for concern and action over fake news content on social media, driven in no small part by accusations of ‘fake news’ among rival politicians. Despite persistent questions about the actual reach of fake news or its efficacy in changing

¹¹⁴⁷ *Paxton* (n 1129) 88-89; see also Crawford (n 1108) 2365-72 (describing the long history of the common carrier doctrine in the common law and its specific applicability to internet service providers)

¹¹⁴⁸ *Att’y Gen. of Fla. v. NetChoice, LLC*, No. ____, *Petition for a Writ of Certiorari* (21 September 2022) (“*Petition*”); see also *NetChoice, LLC v. Moody*, No. 21-12355 (11th Cir., 23 May 2022)

¹¹⁴⁹ *Petition* (n 1148) i

¹¹⁵⁰ *Gonzalez v. Google, LLC*, No. 21-1333, *cert. granted* (03 October 2022)

the minds of voters, copious scholarship has been devoted to describing the fake news phenomenon, often in ways that rely on arbitrary definitions or reify researchers' normative preconceptions. The social media platforms have taken the concerns seriously, adopting new strategies and hiring misinformation 'experts' both in the hopes of improving their services and staving off government intervention. Current public policy responses to its potential election integrity threat have eschewed narrower, elections-oriented approaches in favour of comprehensive regulation for large social media platforms. Though backstopped by notions of foreign or domestic elections interference, the aims of such regulatory regimes are more numerous, drawing in a host of social media behaviours. And because these regimes rely on the social media platforms to accomplish their (sometimes effusive) aims, a thorough consideration of the role of fake news in elections must come to grips with the technical capabilities of the social media platforms more broadly.

As a system for centralizing distribution of privately created content, social media can contribute to a radically more public 'public sphere' only insofar as it allows additional public participation relative to past media systems. The platform infrastructure that enables massively connected social networks is what can promote its public function: the circulation of information from diverse sources and facilitation of public discussion about news events. The connectedness through the platforms—made possible through extensive technical investment and through other advances like mobile phones—distinguishes social media from print, but also makes it more like telegraphy, mail, or broadcasting in the dependence of users upon access to the communication network. We all have printing presses now, but no guarantee our news will be distributed through the distribution channel everyone uses. The way that the social media platforms control access to their centralized network infrastructure is problematic both in terms of the limited capability of the platforms to accurately decide what content should be accessible and the limited answerability of the platforms to users about content determinations. As observed by Branscomb, "If we must learn to cope with bigness, the primary concern of a public philosophy for our communications system should be: How do we democratize the process? How do we involve people in the decision-making?"¹¹⁵¹

How the law handles the platform-centralization of social networks will reflect our public preferences over where to vest authority over the speech and debate transmitted via those networks. In light of a deeper reading of history, the Section 230 advantage of Internet intermediaries over past communications industries is profoundly significant. As other communications technologies developed, the flow of regulatory authority moved from the industrialists who owned the telegraphs to the legislature, from the legislature to the

¹¹⁵¹ Branscomb (n 1097) 121

executive agencies that oversaw the broadcasters, and then, partially, from the public administrators to the cable industry. Today's social media giants were born in a legal milieu that left to the platforms the exclusive power to (electively) accomplish an apparently limitless aim of moderating 'offensive' content.

Faced with a nail sticking up in the form of fake news, policy makers may look in their toolbox for a hammer but find they only have a regulatory scalpel and a market chainsaw. Social and political adaptations require innovation just as much as technologies, and "the careless habit of attributing to mechanical improvements a direct role as instruments of culture and civilization puts a demand upon the machine to which it cannot respond."¹¹⁵² The machine of regulation makes something that isn't useful to the platforms: hierarchical relationships and normative expectations that don't support expanding networks. The machine of markets can produce something that governments may not like: consumer choices that sharply diverge from the consensus view of what's good for the public. But as my investigations have shown, the machine of algorithmic content moderation can't necessarily make what either side wants: diverse, quirky networks of free-thinking social media users that don't occasion conflict or disagreement over electoral politics.

These actualities may even suggest another way out of the dysfunction of social media by leveraging the power of social networking rather than seeking to constrain it—embracing a "pluralism of interests and values" in a single, shared political space.¹¹⁵³ This conception of a more agonistic democracy¹¹⁵⁴ could be furthered by embracing Winston and Winston's critique of the fallacy of 'objective' journalism, which identifies the impossibility of a journalism that can hold itself out as unbiased and error-free¹¹⁵⁵ and argues that abandoning such pretensions in favour of openly partisan (though still informative) journalism would better support the core functions of journalism.¹¹⁵⁶ Dispelling the illusion of 'objective' news would unclothe the exercise of politics in information, and an unfiltered social media space would revive the notional 19th Century parlour or coffee house as a common arena for public conversation among adversaries. The alternative is reification of present social antagonism, each side holding 'genuine' news behind high walls and bitter enmity.

¹¹⁵² Mumford (n 1008) 215

¹¹⁵³ Panu Minkkinen, 'Agonism, Democracy, and Law' in Simon Stern and others, eds *The Oxford Handbook of Law and Humanities* (2019) 426, 438

¹¹⁵⁴ *ibid* 435-39 (discussing Chantal Mouffe's theorization of political adversity as fundamental to democracy *contra* hegemonic practices characterizing 'deliberative' or market democracy)

¹¹⁵⁵ Brian Winston and Matthew Winston, *The Roots of Fake News: Objecting to Objective Journalism* (Routledge 2021) 9-13

¹¹⁵⁶ *ibid* 197-201

The power of the platforms, as with past infrastructure like the telegraphs in the U.S., has invited influence and manipulation, whether sought or unsought, from those with material interests in what the citizenry thinks about political events. Established news media and government agencies alike place high value on maintaining particular narratives in public discourse and suppressing contrary accounts, and accordingly seek to leverage the platforms' central control over access to content in much the same way as they related to one another under broadcast regulation. The individual users of social media lack the institutional gravitas, professional norms, or secret intelligences upon which these entities may rely in making their case to the platforms about certain news content. Such disparities would remain unresolved through approaches like regulation or co-regulation that retain platform-centralized control over content. As a result, social media system dysfunction—failing to provide users with sufficient information for satisfactory civic participation—will likely persist under proposals like the Online Safety Bill. Users will still complain about 'fake news' they don't like, and the platforms may become less able to host a variety of diverse viewpoints.

Approaches that decentralize control over content, on the other hand, can vest power over access in independent structures that individual users can better hold to account. Relying on the courts to determine whether a user has posted false content that interfered with an election leverages a system that has evolved over centuries to be good at establishing facts. The judicial system does this in a public way through a process that is generally viewed as credible and decisive, and through appeals allows its participants to test the rigour and quality of its judgements. But as discussed previously, the challenges for the judiciary in such a role would be many: adequate provisioning for the courts, timeliness of hearings and decisions, and appropriate remedies or penalties for the nebulous harm of someone voting a particular way.

Relying on legislative action to re-align the structure of the social media system could provide creative options that get ahead of *ex post* determinations about fake news while expressing clearer normative aims than regulation. Universal access requirements would drop existing barriers to the circulation of information, while interoperability requirements would empower users to make informed choices about the software they use to filter the content they receive. The challenge for legislators following such approaches will be to overcome the interests entrenched in the status quo—particularly where the platforms argue against the constitutionality of applying long settled doctrines like common carriage.

No conception of social media is without trade-offs, and society inevitably must decide the order of its values to make clear what should be done about news and views on

social media. The obligation of the law is not to protect the current business model of private industry but to promote what works: time-tested values like free expression and the right to hear and receive information that underpin our civilization. We favour other priorities at our own peril and should do so with full understanding of what we exchange for peace of mind about what strangers are saying to each other on Twitter. Law concerning the platforms has the opportunity to set clear aims for the social media system by declaring that everyone in society can participate via this technology and that it is not the place of concentrated private power to decide whether what someone has to say is acceptable. “The ultimate solution is to develop citizens who are capable of comprehending the intricacies of the system. There is something wrong with a system that purports to be democratic but excludes citizens from participating in institutions that influence their daily lives as pervasively as the media.”¹¹⁵⁷ The relation of social media to broader society cannot be improved by pundits and politicians screaming ‘fake news’ until the platforms plug our ears.

Key Takeaways

- Perspectives on what’s to be done about fake news influence reflect a myriad of histories, cultural influences, values, and interests that vary across regions—not all societies will agree on which aspects should take precedence over others
- My research reveals deeper challenges to resolving questions of media influence through automated content moderation, a process which figures centrally in the status quo operations of social media platforms as well as nascent regulatory efforts
- Going beyond modes of responding to media influence that have evolved in 20th Century Britain and America (including both regulation *and* deregulation), historical responses to systemic problems of content inauthenticity and distribution concentration suggest alternatives for improving the civic function of social media
- Decisions about where to address social media system dysfunction are also decisions about where to vest substantial power over public discourse: authorizing private platforms or state regulators to police news content through algorithmic moderation may carry certain risks to free expression, while relying on courts or the public itself to discern news authenticity may carry other social, financial, or political costs
- Such choices require prioritizing particular values and interests over others, and in democratic societies, should proceed only with frank, informed, and open debate

¹¹⁵⁷ Branscomb (n 1097) 129-30

about which values we may prefer and what trade-offs we can accept—especially when they bear on fundamental rights of free expression and political representation

Appendix A

Test Instrument

[Note: Qualtrics tags and procedural information are shown in gray or shaded and were not visible to participants.]

Social Media News Content

Start of Block: Informed Consent

Informed Consent

Election 2020: Can Americans build effective political awareness over social media?

Name of Researcher: Bradley Foster

Dear potential participants,

Thank you for your interest in this project on the role of social media in the U.S. electoral process. The following 15-20 minute survey is part of a research project by the above researcher for his PhD. You may contact us to ask any questions if you would like more information before taking part in the research (see contact details below).

In this study, we will ask you about yourself, your politics, and your impressions on a series of news stories as they would appear on social media, as well as some general political opinions. With your help, we can better understand the role of social media in the national conversation.

Your participation is voluntary, and you are free to withdraw at any time without giving any reason and without it affecting you at all. The last point of withdrawal is at the end of the survey when we will ask for your consent to include your responses. If you choose to withdraw early, you can contact the researchers by e-mail to request a debriefing if desired.

Your personal information will not be shared outside of the research team or published in the final report(s) from this study. All information which you provide during the study will be stored in accordance with the 2018 General Data Protection Regulation and kept strictly confidential.

By answering the questions that follow, you affirm that you have read and understood the above information, you consent to take part in the study and for your data to be used in academic research and for publications, and that you are 18 years of age or older.

We are interested in your genuine responses so please take the questions at face value and give your personal opinion. If you have any questions about this study please contact the researchers.

Researcher contact details:

Bradley Foster: b.foster@uea.ac.uk

Charles Seger: c.seger@uea.ac.uk

If you have any worries or concerns about this research, contact the School of Psychology:

Ethics Committee: ethics.psychology@uea.ac.uk; Phone 01603 597146

Head of School Professor Kenny Coventry: k.coventry@uea.ac.uk; Phone 01603 597145

- ☐ I consent to participate in this study
- ☐ I do NOT consent to participate in this study

Skip To: End of Survey If Informed Consent Election 2020: Can Americans build effective political awareness over social med... = I do NOT consent to participate in this study

End of Block: Informed Consent

Start of Block: Interlude1

The following demographic questions give us some insight into what kinds of people are taking part in the study.

End of Block: Interlude1

Start of Block: Demographics

What is your age?

What is the annual gross income (before taxes) of yourself and any family who lives with you?

- ☐ Less than \$62,000
- ☐ \$62,000 or greater

What is your sex/gender?

- ☐ Male
- ☐ Female
- ☐ Other

What is your race/ethnicity?

- ☐ White or European American
- ☐ Hispanic, Latin or South American
- ☐ Black or African American
- ☐ American Indian or Alaskan Native
- ☐ East Asian
- ☐ Native Hawaiian or other Pacific Islander
- ☐ Middle Eastern
- ☐ South Asian
- ☐ Other (please specify) _____
-

Which state or U.S. territory do you live in?

▼ Alabama ... Puerto Rico or U.S. Virgin Islands

Are you a registered voter?

- ☐ Yes
- ☐ No
- ☐ Not sure
-

Display This Question:

If Even if you don't consider yourself a Republican or Democrat, if you had to pick between the two,... = Democrat

Please rate the following statements:

	Do not agree at all	Mostly disagree	Disagree somewhat	Neither agree nor disagree	Agree somewhat	Mostly agree	Agree completely
I see myself as a Democrat	☐	☐	☐	☐	☐	☐	☐
I am pleased to be a Democrat	☐	☐	☐	☐	☐	☐	☐
I feel strong ties with Democrats	☐	☐	☐	☐	☐	☐	☐
I identify with other Democrats	☐	☐	☐	☐	☐	☐	☐

End of Block: Demographics

Start of Block: Interlude2

The following news stories come from a popular social media site. Please read each item carefully and respond to the accompanying prompts just by giving your first impressions of what you read.

End of Block: Interlude2

Start of Block: N1DemTrue

Just friends? Private flights, luxury vacations, and campaign cash

Democratic Senator Robert Menendez of New Jersey avoided federal corruption charges due to mistrial after jurors had heard nine weeks of evidence about lavish gifts to the politician.

Which of the following **best** describes how you felt after reading this story?

- ☐ Hopeful
- ☐ Irritated
- ☐ Satisfied
- ☐ Guilty
- ☐ Disgusted
- ☐ Uneasy

Thinking about what you read...

	Definitely would not 1	2	3	4	5	6	Definitely would 7
Would you share this news story on social media?	☐	☐	☐	☐	☐	☐	☐
Would you flag or report this story as fake on social media?	☐	☐	☐	☐	☐	☐	☐

End of Block: N1DemTrue

Start of Block: N2DemTrue

Civil rights advocate retires under cloud of controversy

John Conyers, Jr., a Democrat who served as U.S. Representative for Michigan, has announced his retirement after the publication of documents revealing he had settled a sexual harassment case with a former employee.

Which of the following **best** describes how you felt after reading this story?

- ☐ Hopeful
- ☐ Irritated
- ☐ Satisfied
- ☐ Guilty
- ☐ Disgusted
- ☐ Uneasy

Thinking about what you read...

	Definitely would not 1	2	3	4	5	6	Definitely would 7
Would you share this news story on social media?	☐	☐	☐	☐	☐	☐	☐
Would you flag or report this story as fake on social media?	☐	☐	☐	☐	☐	☐	☐

End of Block: N2DemTrue

Start of Block: N3DemTrue

Former Congresswoman ran fake charity, spent over \$300,000 on football tickets, Beyoncé concert

Corrine Brown, a former longtime Democratic U.S. Representative from Florida, was sentenced to five years in prison for operating a fraudulent charity.

Which of the following **best** describes how you felt after reading this story?

- ☐ Hopeful
- ☐ Irritated
- ☐ Satisfied
- ☐ Guilty
- ☐ Disgusted
- ☐ Uneasy

Thinking about what you read...

	Definitely would not 1	2	3	4	5	6	Definitely would 7
Would you share this news story on social media?	☐	☐	☐	☐	☐	☐	☐
Would you flag or report this story as fake on social media?	☐	☐	☐	☐	☐	☐	☐

End of Block: N3DemTrue

Start of Block: N4DemTrue

Youthful Indiscretions? 19-year old candidate’s future uncertain

Aaron Coleman, Democratic candidate for the Kansas Legislature, wins primary contest but has admitted to sending revenge porn and bullying girls online when he was in middle school.

Which of the following **best** describes how you felt after reading this story?

- ☐ Hopeful
- ☐ Irritated
- ☐ Satisfied
- ☐ Guilty
- ☐ Disgusted
- ☐ Uneasy

Thinking about what you read...

	Definitely would not 1	2	3	4	5	6	Definitely would 7
Would you share this news story on social media?	☐	☐	☐	☐	☐	☐	☐
Would you flag or report this story as fake on social media?	☐	☐	☐	☐	☐	☐	☐

End of Block: N4DemTrue

Start of Block: N5DemTrue

Rising star crashes and burns

U.S. Representative Katie Hill of California, a first-term Democrat, has denied the allegation of a sexual relationship with a staff member, which would violate House rules.

Which of the following **best** describes how you felt after reading this story?

- ☐ Hopeful
- ☐ Irritated
- ☐ Satisfied
- ☐ Guilty
- ☐ Disgusted
- ☐ Uneasy

Thinking about what you read...

	Definitely would not 1	2	3	4	5	6	Definitely would 7
Would you share this news story on social media?	☐	☐	☐	☐	☐	☐	☐
Would you flag or report this story as fake on social media?	☐	☐	☐	☐	☐	☐	☐

End of Block: N5DemTrue

Start of Block: N6DemFake

New spending on research & development, but who benefits?

U.S. Representative Leanne Rushton, a California Democrat, is under investigation for buying stock in a genetics research firm about to be awarded a federal contract – before its public announcement.

Which of the following **best** describes how you felt after reading this story?

- ☐ Hopeful
- ☐ Irritated
- ☐ Satisfied
- ☐ Guilty
- ☐ Disgusted
- ☐ Uneasy

Thinking about what you read...

	Definitely would not 1	2	3	4	5	6	Definitely would 7
Would you share this news story on social media?	☐	☐	☐	☐	☐	☐	☐
Would you flag or report this story as fake on social media?	☐	☐	☐	☐	☐	☐	☐

End of Block: N6DemFake

Start of Block: N7DemFake

Big Kahuna’s nephew in hot water over competency concerns

Allan Lono, 29, has denied reports he benefitted from family, Democratic Party connections to get his position as a Hawai’i state official, despite calls for his resignation.

Which of the following **best** describes how you felt after reading this story?

- ☐ Hopeful
- ☐ Irritated
- ☐ Satisfied
- ☐ Guilty
- ☐ Disgusted
- ☐ Uneasy

Thinking about what you read...

	Definitely would not 1	2	3	4	5	6	Definitely would 7
Would you share this news story on social media?	☐	☐	☐	☐	☐	☐	☐
Would you flag or report this story as fake on social media?	☐	☐	☐	☐	☐	☐	☐

End of Block: N7DemFake

Start of Block: N8DemFake

Quid pro quo uncovered by animal rights activists

Gerry Pressley, a Department of Agriculture official appointed by President Obama, faces corruption charges over gifts received in exchange for favorable treatment of certain federal licensing renewals.

Which of the following **best** describes how you felt after reading this story?

- ☐ Hopeful
- ☐ Irritated
- ☐ Satisfied
- ☐ Guilty
- ☐ Disgusted
- ☐ Uneasy

Thinking about what you read...

	Definitely would not 1	2	3	4	5	6	Definitely would 7
Would you share this news story on social media?	☐	☐	☐	☐	☐	☐	☐
Would you flag or report this story as fake on social media?	☐	☐	☐	☐	☐	☐	☐

End of Block: N8DemFake

Start of Block: N9DemFake

How one candidate tried to cash out before the race was over

Jeff Anderson, Democratic candidate in a Pennsylvania congressional race has apologized after disclosures revealed the misuse of campaign cash for personal expenses, lavish steakhouse dinners.

Which of the following **best** describes how you felt after reading this story?

- ☐ Hopeful
- ☐ Irritated
- ☐ Satisfied
- ☐ Guilty
- ☐ Disgusted
- ☐ Uneasy

Thinking about what you read...

	Definitely would not 1	2	3	4	5	6	Definitely would 7
Would you share this news story on social media?	☐	☐	☐	☐	☐	☐	☐
Would you flag or report this story as fake on social media?	☐	☐	☐	☐	☐	☐	☐

End of Block: N9DemFake

Start of Block: N10DemFake

Blocked and reported

A Democratic Chicago alderman has denied allegations his wife’s tech non-profit wrongly benefited from a million dollar block grant directed to his ward.

Which of the following **best** describes how you felt after reading this story?

- ☐ Hopeful
- ☐ Irritated
- ☐ Satisfied
- ☐ Guilty
- ☐ Disgusted
- ☐ Uneasy

Thinking about what you read...

	Definitely would not 1	2	3	4	5	6	Definitely would 7
Would you share this news story on social media?	☐	☐	☐	☐	☐	☐	☐
Would you flag or report this story as fake on social media?	☐	☐	☐	☐	☐	☐	☐

End of Block: N10DemFake

Start of Block: N11RepTrue

Explicit photos kick up storm on Capitol Hill

Texas Republican Congressman Joe Barton has apologized for sexually explicit photographs he exchanged with a woman after one of the photographs appeared on an anonymous Twitter account.

Which of the following **best** describes how you felt after reading this story?

- ☐ Hopeful
- ☐ Irritated
- ☐ Satisfied
- ☐ Guilty
- ☐ Disgusted
- ☐ Uneasy

Thinking about what you read...

	Definitely would not 1	2	3	4	5	6	Definitely would 7
Would you share this news story on social media?	☐	☐	☐	☐	☐	☐	☐
Would you flag or report this story as fake on social media?	☐	☐	☐	☐	☐	☐	☐

End of Block: N11RepTrue

Start of Block: N12RepTrue

Whatcha gonna do, brother? Politician body-slams reporter

U.S. Representative Greg Gianforte, a Montana Republican, was sentenced to community service and anger management classes after body-slaming a reporter the night before he won his seat.

Which of the following **best** describes how you felt after reading this story?

- ☐ Hopeful
- ☐ Irritated
- ☐ Satisfied
- ☐ Guilty
- ☐ Disgusted
- ☐ Uneasy

Thinking about what you read...

	Definitely would not 1	2	3	4	5	6	Definitely would 7
Would you share this news story on social media?	☐	☐	☐	☐	☐	☐	☐
Would you flag or report this story as fake on social media?	☐	☐	☐	☐	☐	☐	☐

End of Block: N12RepTrue

Start of Block: N13RepTrue

Forget picnic baskets – politician tries to make off with the bag

Republican U.S. Representative Chris Collins of New York has pleaded not guilty to insider trading charges alleging he helped his son sell off stocks on a tip received during the Congressional Picnic.

Which of the following **best** describes how you felt after reading this story?

- ☐ Hopeful
- ☐ Irritated
- ☐ Satisfied
- ☐ Guilty
- ☐ Disgusted
- ☐ Uneasy

Thinking about what you read...

	Definitely would not 1	2	3	4	5	6	Definitely would 7
Would you share this news story on social media?	☐	☐	☐	☐	☐	☐	☐
Would you flag or report this story as fake on social media?	☐	☐	☐	☐	☐	☐	☐

End of Block: N13RepTrue

Start of Block: N14RepTrue

“Jekyll and Hyde” politician lived extravagant double life

U.S. Representative Duncan Hunter, a California Republican, has been indicted for spending almost a quarter of a million dollars of campaign money on personal expenses, including parties and bar tabs.

Which of the following **best** describes how you felt after reading this story?

- ☐ Hopeful
- ☐ Irritated
- ☐ Satisfied
- ☐ Guilty
- ☐ Disgusted
- ☐ Uneasy

Thinking about what you read...

	Definitely would not 1	2	3	4	5	6	Definitely would 7
Would you share this news story on social media?	☐	☐	☐	☐	☐	☐	☐
Would you flag or report this story as fake on social media?	☐	☐	☐	☐	☐	☐	☐

End of Block: N14RepTrue

Start of Block: N15RepTrue

Who says voter fraud never happens anymore?

Steve Watkins, a Kansas Republican, will face charges of voter fraud after reportedly providing false information to law enforcement and voting unlawfully.

Which of the following **best** describes how you felt after reading this story?

- ☐ Hopeful
- ☐ Irritated
- ☐ Satisfied
- ☐ Guilty
- ☐ Disgusted
- ☐ Uneasy

Thinking about what you read...

	Definitely would not 1	2	3	4	5	6	Definitely would 7
Would you share this news story on social media?	☐	☐	☐	☐	☐	☐	☐
Would you flag or report this story as fake on social media?	☐	☐	☐	☐	☐	☐	☐

End of Block: N15RepTrue

Start of Block: N16RepFake

‘No undue influence’ was exerted over contracting decisions

Oregon state senator Barbara Durkheim, a Republican, responded on Thursday to claims her family’s waste removal business stands to profit unfairly from state contracts.

Which of the following **best** describes how you felt after reading this story?

- ☐ Hopeful
- ☐ Irritated
- ☐ Satisfied
- ☐ Guilty
- ☐ Disgusted
- ☐ Uneasy

Thinking about what you read...

	Definitely would not 1	2	3	4	5	6	Definitely would 7
Would you share this news story on social media?	☐	☐	☐	☐	☐	☐	☐
Would you flag or report this story as fake on social media?	☐	☐	☐	☐	☐	☐	☐

End of Block: N16RepFake

Start of Block: N17RepFake

Private Prison Blues: Which lawmakers pushed deal

Alabama Republican state representative Sue Walker faces hometown backlash over funding plan for the construction of a new state prison in her district.

Which of the following **best** describes how you felt after reading this story?

- ☐ Hopeful
- ☐ Irritated
- ☐ Satisfied
- ☐ Guilty
- ☐ Disgusted
- ☐ Uneasy

Thinking about what you read...

	Definitely would not 1	2	3	4	5	6	Definitely would 7
Would you share this news story on social media?	☐	☐	☐	☐	☐	☐	☐
Would you flag or report this story as fake on social media?	☐	☐	☐	☐	☐	☐	☐

End of Block: N17RepFake

Start of Block: N18RepFake

Triggered? Non-profit breached political advocacy restrictions

A Republican candidate for a Texas congressional seat has apologized after the publication of campaign finance records revealed improper donations from a gun rights group.

Which of the following **best** describes how you felt after reading this story?

- ☐ Hopeful
- ☐ Irritated
- ☐ Satisfied
- ☐ Guilty
- ☐ Disgusted
- ☐ Uneasy

Thinking about what you read...

	Definitely would not 1	2	3	4	5	6	Definitely would 7
Would you share this news story on social media?	☐	☐	☐	☐	☐	☐	☐
Would you flag or report this story as fake on social media?	☐	☐	☐	☐	☐	☐	☐

End of Block: N18RepFake

Start of Block: N19RepFake

Swamp gas and the appearance of impropriety

Republican state senator Gregg Thibodeaux has been accused of insider trading after buying stock in two companies about to be awarded natural gas exploration rights in Louisiana.

Which of the following **best** describes how you felt after reading this story?

- ☐ Hopeful
- ☐ Irritated
- ☐ Satisfied
- ☐ Guilty
- ☐ Disgusted
- ☐ Uneasy

Thinking about what you read...

	Definitely would not 1	2	3	4	5	6	Definitely would 7
Would you share this news story on social media?	☐	☐	☐	☐	☐	☐	☐
Would you flag or report this story as fake on social media?	☐	☐	☐	☐	☐	☐	☐

End of Block: N19RepFake

Start of Block: N20RepFake

Peer pressure amidst Congress members over stock trades

U.S. Representative Bill Gulch, an Iowa Republican, is under investigation by the House Committee on Ethics after he sold Monsanto stock ahead of the public release of findings from a closed hearing.

Which of the following **best** describes how you felt after reading this story?

- ☐ Hopeful
- ☐ Irritated
- ☐ Satisfied
- ☐ Guilty
- ☐ Disgusted
- ☐ Uneasy

Thinking about what you read...

	Definitely would not 1	2	3	4	5	6	Definitely would 7
Would you share this news story on social media?	☐	☐	☐	☐	☐	☐	☐
Would you flag or report this story as fake on social media?	☐	☐	☐	☐	☐	☐	☐

End of Block: N20RepFake

Start of Block: Interlude3

The remaining questions will help us understand public opinions about general political issues and social media use.

End of Block: Interlude3

Start of Block: Follow Up

Before I share a story on social media, I usually read...

- ☐ just the headline
 - ☐ a little of the story
 - ☐ at least half the story
 - ☐ most of the story
 - ☐ the entire story
-

To what extent do you think the "official version" of some events could be an attempt by the authorities to hide the truth from the public?

- ☐ Almost never 1
 - ☐ 2
 - ☐ 3
 - ☐ 4
 - ☐ 5
 - ☐ 6
 - ☐ Nearly always 7
-

How important are each of the following to you?

	Not at all important 1	2	3	4	5	6	Very important 7
Balanced budget amendment	☐	☐	☐	☐	☐	☐	☐
Anti-discrimination	☐	☐	☐	☐	☐	☐	☐
School choice	☐	☐	☐	☐	☐	☐	☐
The Affordable Care Act	☐	☐	☐	☐	☐	☐	☐
Securing the border	☐	☐	☐	☐	☐	☐	☐
Relying on facts	☐	☐	☐	☐	☐	☐	☐
Gun control	☐	☐	☐	☐	☐	☐	☐
Fighting climate change	☐	☐	☐	☐	☐	☐	☐
Freedom of speech	☐	☐	☐	☐	☐	☐	☐
Freedom of the press	☐	☐	☐	☐	☐	☐	☐
Religious liberty	☐	☐	☐	☐	☐	☐	☐
Hearing both sides	☐	☐	☐	☐	☐	☐	☐

When you think about the **Democratic Party**, to what extent do you feel each of the following emotions?

	Not at all 1	2	3	4	5	6	Very much 7
Angry	☐	☐	☐	☐	☐	☐	☐
Satisfied	☐	☐	☐	☐	☐	☐	☐
Afraid	☐	☐	☐	☐	☐	☐	☐
Hopeful	☐	☐	☐	☐	☐	☐	☐
Proud	☐	☐	☐	☐	☐	☐	☐
Disgusted	☐	☐	☐	☐	☐	☐	☐
Uneasy	☐	☐	☐	☐	☐	☐	☐
Happy	☐	☐	☐	☐	☐	☐	☐
Grateful	☐	☐	☐	☐	☐	☐	☐
Guilty	☐	☐	☐	☐	☐	☐	☐
Respectful	☐	☐	☐	☐	☐	☐	☐
Irritated	☐	☐	☐	☐	☐	☐	☐

When you think about the **Republican Party**, to what extent do you feel each of the following emotions?

	Not at all 1	2	3	4	5	6	Very much 7
Angry	☐	☐	☐	☐	☐	☐	☐
Satisfied	☐	☐	☐	☐	☐	☐	☐
Afraid	☐	☐	☐	☐	☐	☐	☐
Hopeful	☐	☐	☐	☐	☐	☐	☐
Proud	☐	☐	☐	☐	☐	☐	☐
Disgusted	☐	☐	☐	☐	☐	☐	☐
Uneasy	☐	☐	☐	☐	☐	☐	☐
Happy	☐	☐	☐	☐	☐	☐	☐
Grateful	☐	☐	☐	☐	☐	☐	☐
Guilty	☐	☐	☐	☐	☐	☐	☐
Respectful	☐	☐	☐	☐	☐	☐	☐
Irritated	☐	☐	☐	☐	☐	☐	☐

To what extent do you think your friends and contacts on social media have the same beliefs as you?

Please indicate on the scale below what percentage you think best represents your similarity, from 0% (completely different beliefs from yours) to 100% (completely same beliefs as yours).

Completely different

Completely same

0102030405060708090100

As a percentage:

If the election for president were going to be held right now and the Democratic nominee was Joe Biden and the Republican nominee was Donald Trump, who would you vote for?

- ☐ Donald Trump
- ☐ Joe Biden
- ☐ Other
- ☐ Not sure
- ☐ I would not vote

In what year did an unidentified flying object allegedly crash in Roswell, New Mexico? We just have to make sure you've been reading carefully, so please enter 1947 in the box below.

Start of Block: Debriefing

Debriefing

Election 2020: Can Americans build effective political awareness over social media?

Thank you for participating in this study. Your time and efforts are much appreciated.

Your data will be anonymized and used as group data that is useful for understanding the influence of political engagement on decision making about news content on social media.

We know that relations between groups can shape how we understand present events, such as when fans of opposing basketball teams disagree over the wisdom of a referee's foul call. Our goal with this study is to extend our knowledge of intergroup relations by examining how our thoughts and feelings about our political groups shape and motivate our decisions about information we encounter over social media.

Some of the news stories featured in this study were created by the researchers for the purposes of the experiment. If you have any questions about the content of the news stories you read, please feel free to research those topics further after you complete this debriefing.

If you have any questions regarding this study, please feel free to ask or contact the researchers of this study, now or at a later date.

Sometimes people will not want us to use their responses because, for example, they feel they didn't pay enough attention to the questions or they don't want to take part in the research anymore. **Should we use your anonymous questionnaire responses?**

This will not affect your compensation either way.

- ☐ Yes, use my responses
- ☐ No, do NOT use my responses

You will need to copy and paste the following number into MTurk/Cloud Research to be compensated for your participation.

Once you've copied the number, please click the next button to finish the questionnaire.

Your random identification number: \${e://Field/Random%20ID}

End of Block: Debriefing

Appendix B

Research Ethics Application

UNIVERSITY OF EAST ANGLIA SCHOOL OF PSYCHOLOGY RESEARCH ETHICS APPLICATION

Section I: Project Details

Project reference:	2020-1283-001997
Project Title: This will be the title through review.(max 100 chars)	Intergroup influences on news content sharing over social media
SONA Title: This will be the exact title you use on SONA.(max 100 chars)	N/A
Project Funder:	UEA
R project code:	N/A
Project officer:	N/A
Can you confirm that financial gain will not be indicated in the title of your advert (if SONA credits are being awarded)	yes
Start Date:	01/10/2020
End Date:	01/10/2021
Brief Summary of the Project (not more than 100 words):	This project aims to provide evidence of the mediating role of political group identification in decision making about sharing and flagging news information over social media. We will ask participants about themselves, their political affiliation, and their impressions on a series of true and fake news stories; as well as for evaluations of their political ingroups and outgroups. We expect the data collected will highlight the role of intergroup relations in sharing/reporting decisions, which may be critically unexamined in regulatory or institutional policy determinations regarding news content on social media.
If a standard research protocol with existing ethics approval is being used please give the title of the protocol and the reference number.	N/A

Has your study already received a Research Ethics Review, or has been reviewed in full by another institution because it is part of a larger study? If yes, you are required to upload the approval documentation as an appendix.	No
---	----

Section II: Applicant Details

Name of applicant:	Bradley Foster
Name or researcher who is gathering the data:	Bradley Foster
Supervisor:	Charlie Seger
School:	Psychology
Email address:	b.foster@uea.ac.uk
Telephone number:	07947 214255

Section III: For Taught Students Only (UG and PG)

Course:	N/A
Department:	N/A
Module:	N/A
Module number:	N/A
Module leader's name:	N/A

Section IV: Research Checklist

Will the study involve recruitment of patients through the NHS or Social Care, or the use of NHS patient data or premises and/or equipment? All research involving patient data must be reviewed by the NHS NRES (Recruitment of NHS staff or use of NHS data/equipment can go through the psychology ethics committee.)	No
---	----

Does the study involve participants age 16 or over who are unable to give informed consent (e.g. people with learning disabilities: see Mental Capacity Act 2005)? All research that falls under the auspices of the MCA must be reviewed by NHS NRES	No
Will tissue samples (including blood) be obtained from participants? All research involving human tissue must be reviewed by the UEA Faculty of Medicine and Health Sciences Research Ethics Committee	No
Would you like to continue and complete the full ethics approval checklist and review even though you have answered YES to some of the above?	No

If you have answered 'Yes' to any of the questions above you will need to submit your research for ethical approval to the appropriate external body. See the [UEA Research Governance Guidance for Researchers and Supervisors](#) **Send the completed and signed Checklist to the Deputy Chair for registration.**

Once ethical approval is granted by the external body, a copy of the approval should be sent to your School Research Ethics Officer (SCI) or Faculty Research Ethics Administrator for their records.

Section IV: Research Checklist

Does the research involve animals?	No
Does the research involve vulnerable groups (children, those with cognitive impairment, or those in unequal relationships e.g. your own students in class)?	No

Will the study require the co- operation of a gatekeeper/subject panel for initial access to the groups or individuals to be recruited (e.g. students at school, members of self-help group, residents of Nursing home, prisoners)?	No
Will it be necessary for participants to take part in the study without their knowledge and consent at the time (e.g. covert observation of people in non-public places)?	No
Will deception be used?	No
Will the study involve discussion of sensitive topics (e.g. sexual activity, drug use, ethnicity, political behaviour) or involve elite interviews?	Yes
Will the research involve access to records of personal/ sensitive/ confidential information , or involve commercial confidentiality/national security?	No
Are drugs, placebos or other substances (e.g. food substances , vitamins) to be administered to the study participants or will the study involve invasive, intrusive or potentially harmful procedures of any kind?	No
Is pain or more than mild discomfort likely to result from the study?	No
Could the study induce psychological stress or anxiety or cause harm or negative consequences beyond the risks encountered in normal life?	No
Will the study involve prolonged or repetitive testing beyond a regular laboratory experiment?	No

Will the research involve administrative or secure data that requires permission from the appropriate authorities before use?	No
Is there a possibility that the safety of the researcher may be in question (e.g. in international research: locally employed research assistants)?	No
Does the research involve members of the public in a research capacity (that is, the participants themselves will be carrying out research)?	No
Will the research take place outside the UK?	Yes
Will the research expose respondents to the internet or other visual/vocal methods where respondents may be identified ?	Yes
Will research involve the sharing of data or confidential information beyond the initial consent given (e.g. secondary use of data)?	No
Will financial inducements (other than reasonable expenses and compensation for time as in regular laboratory experiments) be offered to participants?	No
Will your research involve investigation of or engagement with terrorist or violent extremist groups? Please provide a full explanation if the answer is yes.	No
Does your research have environmental implications? Please refer to the University's Research Ethics Guidance Note: Research with a Potential Impact on the Environment for further details.	No

Is there potential for your research to affect cultural objects?	No
--	----

If you have answered Yes to ANY of questions on this page, please explain your YES-answers briefly:(max 300 words)

Some sensitive topics are mentioned in several of the “news items” to be evaluated by participants. Participants will have to read and reflect on a series of news headlines and lead-in sentences covering a variety of public interest stories about politicians that include, in a few instances, references to sexual assault and other crimes. Participants will also appraise the importance of a set of political topics, including gun violence and immigration, and give their affective evaluation of their political ingroup and outgroup. However, no sensitive issue is discussed further beyond their mere presentation as part of these devices. I will use participants only in the United States. The study will be conducted over the Internet via CloudResearch (formerly TurkPrime/MTurk) and any user-specific data made anonymous upon collection.

The issues highlighted above should be considered carefully when completing the full ethical review form which follows.

Section IV: Research Checklist

Is this a project funded by a research council such as the ESRC? Y (full committee)/N –see below	No
Is this a project which is highly sensitive ethically? Y (full committee)/N see below.	No
Is this a staff or postgraduate research project? Y (2 reviewers)/N – see below.	Yes
Is this an undergraduate project on a potentially vulnerable population (including participants under 18 years old)? Y (2 reviewers)/N – see below	No
Is this an undergraduate project which has some sensitive issues and the views of a second reviewer would be beneficial? Y (2 reviewers)/N – see below	No
Is this an undergraduate project in which none of the above apply? Y (1 reviewer)	No

Methods

Background and issues, aims, design (e.g.interview, experimental, observational, survey), research questions / hypotheses (2-300 words):	My thesis will identify crucial considerations which a legal analysis may miss in crafting regulatory responses to the “fake news” problem on social media. Psychology has already made important contributions to jurisprudence generally by explaining some of the ways by which mental processes constrain the evaluation of evidence and the exercise of judgment. Social psychology in particular influences group level interactions with significant legal implications, such as in the public exercise of free expression and the determination of its appropriate restriction, because of the impact of social identification on human judgment and behaviour. My research will highlight the role of social media users' group affiliations in their construal of a particular post as “fake news” or not. The design of this psychological research will be an online, survey-based instrument to collect data on the participants’ political identification and to examine whether participants will decide to report or categorize true news as fake, or share fake news as true. The data generated will be used to test whether participant’s decisions about reporting and sharing vary with the strength of their partisanship. I also want to examine the extent to which participants share certain group objectives or ideals and test whether that influences the share/report decisions.
How many participants do you intend to include in the study? (numeric value only please)	400
What are the characteristics of the participants? (Please list all inclusion and exclusion criteria)(max 300 words)	Participants will be aged 18+ and U.S.-based. No other demographic criteria will affect inclusion or exclusion. Ultimately, participants will be characterised by their strength of identification with either of the two major U.S. political parties, including if that identification is weak. Participants may be excluded from analyses if they do not answer all the questions or fail an “attention” item.
What is the process of recruitment, how will participants be approached and invited to take part?(max 300 words)	Recruitment will take place on Cloud Research’s platform. Potential participants will see an advertisement and will have the opportunity to take part in exchange for appropriate compensation (~\$1.50). Participation should take about 15 to 20 minutes. The following survey explanation will be provided: How to make a difference in the Election: Can Americans build effective political awareness over social media? In this study, we will ask you about yourself, your politics, and your impressions on a series of news stories as they would appear on social media, as well as some general political opinions. With your help, we can better understand the role of social media in the election process. The survey should take around 15-20 minutes and you will be compensated for its completion.
Will external organisations/people’s consent be required?	No

If Yes please detail:(max 300 words)	N/A
Is the planned sample size achievable and appropriate for meaningful data analysis?	Yes
Is this research taking place via the internet/post?	Yes
If the study is conducted via the internet have you included safeguards to ensure that participants are not vulnerable or underage?	Yes
What are those safeguards? (max 300 words)	Underage participants should not be able to sign up for Cloud Research and thus should not be able to access the study. They can also be excluded from the data set if they report an age under 18 on the Age item in the demographics section of the study. Any other potential participants should have ample opportunity to evaluate the study based on the recruitment invitation and informed consent document and then decide whether to take part and are free to withdraw at any time before the end of the study.
If data is being gathered via the internet are you gathering IP addresses?	No
If yes are you ensuring that participants explicitly consent to this?	N/A
Are you using the standard <u>School of Psychology guidelines</u> for participant reimbursement (credits for SONA or payment for funded studies)?	No
If not why not?(max 300 words)	Rather than the standard online participant compensation scheme ("prize fund"), my study will make use of Cloud Research's compensation scheme to provide direct and appropriate remuneration to all participants. I will not be collecting email addresses or other personally identifiable information about participants, so contacting them after the study for a prize fund draw would not be feasible.
Is your recruitment process non- coercive and is it clear there are no consequences for non-participation?	Yes

Please outline what the participants will experience including what measures, materials or apparatus will you use? (Please give details and include copies of questionnaires, interview schedules, experimental stimuli etc. Be mindful that not all research requires asking personal and sensitive questions and this should be considered when deciding on measures) (max 300 words)	Please see Appendix for full copy. Participants will be provided an explanation of the study and a consent form. Using an online questionnaire, consenting participants will be asked about themselves to measure demographics and to assess them on a political identification scale. Participants will then be presented with a series of news headlines and ledes followed by two questions: 1) how likely they would be to share the story; and 2) how likely they would be to report the story as fake. Participants will also have an array of emotional responses from which to pick describing how they felt about the story. The news stories cover public interest events involving Republican and Democratic politicians. Topics include fraud, nepotism, insider trading, sexual harassment, and corruption. Half of the stories are fictitious, but no indication of this is provided to participants. Participants will also be asked to complete a series of follow up questions, including • an assessment of a specific online behaviour • a single item scale on conspiracism • an “importance” inventory of political objectives/ideals • an affective evaluation of the two major U.S. political parties. Finally, we will ask participants for their voting intention in the upcoming election and administer an attention item to make sure they’ve been participating earnestly. Participants will read a short debrief about the aims of the study and be given a final opportunity to withdraw their data.
--	---

Informed consent and briefing

Is informed consent to be obtained from participants?	Yes
Will you append a copy of the invitation letter/advert?	Yes
Will you append a copy of the participant information sheets?	No
Is your participant information sheet based on the official template?	No
If not why not?(max 300 words)	As this is an online study, I will use the “consent for questionnaires” template, adapted to cover the particulars of my project. Accordingly, I won’t be using separate consent and information forms, as the consent for questionnaires form covers the key points of the usual participant information sheet.
Will you append a copy of the consent form?	Yes

Will participants be explicitly informed of what the researcher's role/status is?	Yes
Will participants be told of the use to which data will be put (e.g., research publications, teaching purposes, media publication)?	Yes

Right of withdrawal

When is the last point of withdrawal?(max 300 words)	Participants can withdraw at any point throughout the study by simply stopping the completion of the survey. Additionally, at the end of the study, they will be asked if they would still want their data to be used. They will be informed that they will not be able to withdraw their data after they complete the survey due to its anonymous nature.
Is this clear from your consent/participant information form?	Yes
How will you deal with anonymity issues for late withdrawal (e.g. use of participant codes)?(max 300 words)	Data anonymity at collection means late withdrawal will be unavailable.
Are participants given a genuine, unpressured opportunity to withdraw?	Yes
If NO, explain why not:(max 300 words)	NA

Debriefing

Will the participants be debriefed? (Please append verbal or written text)	Yes
If YES, how will they be debriefed (e.g., verbally, debriefing sheet; give details or attach the debriefing information to this form) or if NO, why not?(max 300 words)	A written debriefing sheet will be shown to participants at the end of the survey. Please see Appendix for a copy.
Does the debrief offer Sources of Support where relevant?	No
Does it offer an easily understandable lay explanation of the	Yes

research?	
Does it contain contact details for the researcher and ethics Committee?	No
If the study is being completed via the internet remember participants may exit part way through without receiving the debrief - does the design and or initial consent information take this into account?	Yes
Are you using the debrief template?	Yes

Confidentiality

Will you meet the participants?	No
Will the participants sign anything?	No
Will IP addresses be collected?	No
If none of the above are answered 'yes' will the data be gathered anonymously?	Yes
If NO, how will you protect the identity of your participants and ensure that any personal information you receive will be kept confidential?(max 300 words)	N/A
Will you remove identifying information from the data and, if necessary, replace it with ID numbers or pseudonyms?	Yes
Will you store data securely (e.g., in a locked filing cabinet or password-protected electronic file)?	Yes
Are you storing contact details such as email addresses separately to responses?	No

Risk assessment

What inconveniences might participants experience?(max 300 words)	Time and attention. Sensitivity: Participants will need to read on-screen brief news items covering politicians' bad conduct and consider their responses. Some topics may be controversial or upsetting, as discussed previously.
What steps will you take to minimize these?(max 300 words)	While the survey is intended to be completed in a single, short session, participants are of course free to take breaks if needed. None of the questions are any more invasive than participants might encounter in an ordinary conversation about politics or the election. While some of the news story topics such as sexual harassment may be sensitive for some participants, their presentation in the study is meant to approximate the sort of news stories any person might see shared over social media and contain no details of any sort. The response prompts encourage participants' first impressions of what they read. Participants are reminded that they can research any of the news stories they want to know more about after completing the survey. Survey participants will be compensated for their time.
Will involvement in the research put participants at risk of physical or psychological harm, distress or discomfort greater than that encountered in their everyday lives?	No
If YES, describe the nature of the risk and the steps you will take to minimise it(max 300 words):	N/A
Will you complete the researcher safety checklist, and will you review the researcher safety checklist for each member of the research team (see appendix)? If the details of the research are not known at this time, then details from the checklist should be logged with the committee prior to collecting data.	No
Is a risk assessment necessary?	No
Does involvement in the research put you at risk of physical or psychological harm, distress or discomfort greater than that encountered in your everyday life?	No

If YES, describe the nature of the risk and the steps you will take to minimise it: (See Researcher Safety Policy).(max 300 words)	N/A
--	-----

Other permissions and clearances

Is ethical clearance required from any other ethics committee?	No
If YES, please give the name and address of the organisation:	N/A
Has such ethical clearance been obtained yet? If YES, please attach a copy of the ethical approval letter	No

Please note that it is your responsibility to follow the University of East Anglia Research Ethics Policy, Principles and Procedures and any relevant academic or professional guidelines in the conduct of your study. **This includes providing appropriate information sheets and consent forms, following appropriate recruitment policies, and ensuring confidentiality in the storage and use of data.**

Any significant change in the question, design or conduct over the course of the research should be notified to the Research Ethics Committee and may require a new application for ethics approval.

Please enter your name and the date to confirm that you are the applicant and have read and understood the above:

Name:	Bradley Foster	Date: 09/09/2020
-------	----------------	------------------

Appendix C

List of Tables

Table C1 – Comparisons Between Participant Groups by Additional Measures	257
Table C2 – Kendall Rank Correlation Table: Additional Measurements	258
Table C3 – Voting Intention Frequency by Party Grouping	259
Table C4 – Kendall Rank Correlation Table: Demographic and Test Variables	260
Table C5 – Group separated weighted least squares regression analyses predicting willingness to share fake outgroup stories from demographic factors and group identification scores	261
Table C6 – Group separated weighted least squares regression analyses predicting willingness to flag genuine ingroup stories as fake from demographic factors and group identification scores	262
Table C7 – Emotional Reaction Frequency by News Item Category	265
Table C8 – Comparisons Between Participant Groups by Political Issues	267
Table C9 – Kendall Rank Correlation Table: Additional Public Discourse Issue	268
Table C10 – Group separated regression analyses predicting willingness to share fake outgroup stories from demographic factors, group identification scores, and importance of relying on facts	269
Table C11 – Sex/gender separated regression analyses predicting willingness to share fake outgroup stories from age, party affiliation, and group identification scores	271
Table C12 – Sex/gender separated regression analyses predicting willingness to flag genuine ingroup stories from age, party affiliation, and group identification scores	272

Additional Measurements

As part of my experiment design, I determined to include a number of additional measurements in the test instrument, some for exploratory purposes and others to assess construct validity—checking to make sure the measures I was using were capturing the specific phenomena I wanted to study regarding political news. Accordingly, I had my participants complete a series of supplemental questions to conclude the testing portion of the experiment. These follow-up items also allowed the participants to describe themselves in ways that aren't captured directly by the test items, enriching the picture of how the participants interact with and think about news stories online.

I included an item asking participants to reflect on their typical reading habits before sharing a news story to provide another reference point for general social media behaviour. The response scale provided the following options: 1) just the headline, 2) a little of the story, 3) at least half the story, 4) most of the story, or 5) the entire story.¹¹⁵⁸ Asking participants whether they usually read news posts before sharing is a bit of a loaded question after having them go through twenty ‘minimal information’ sharing decisions—there is of course the possibility that responses to this item could be coloured by participants’ desire to cast themselves in a more socially responsible light. Likewise, these self reports of reading habits may or may not be accurate to what people actually do on social media. Nevertheless, the relationship between actually reading and later sharing a story is an interesting phenomenon in its own right, and some content moderation strategies like Twitter’s read-before-retweeting suggestion¹¹⁵⁹ point to how seriously social media stakeholders are taking it. My participants’ self-assessments did not suggest that reading-before-sharing behaviour is group-specific, as shown in **Table C1**.

I also included a single item measure of “conspiracism” adapted from Lantian and others¹¹⁶⁰ to collect participants’ general sentiments on conspiracy beliefs. Rather than ask participants how “completely” true or false they thought an affirmative statement about the reliability of official accounts of events was, I asked them to consider the extent to which the “official version” of some events could be an attempt by the authorities to hide the truth from the public, on a scale of one to seven anchored at point one by “Almost never” and at point seven by “Nearly always.”¹¹⁶¹ A comparison between the groups is shown in **Table C1**. This was undertaken in keeping with my reasoning on what makes a news story true or false as discussed in Part Four; I wanted to assess whether my participants had some sense of the distance and barriers between news events and received accounts that Lippmann described.¹¹⁶² The difference between the groups could reflect perceptions of a ‘liberal’ media bias¹¹⁶³ among my Republican participants, but both groups tended to rate this item on the higher end of the scale.¹¹⁶⁴

¹¹⁵⁸ Mean = 4.19, Standard Deviation = 1.099

¹¹⁵⁹ Michael Kan, ‘Twitter: Our “Read Before You Retweet” Function Actually Works’ (*PCMag.com*, 25 September 2020) <<https://www.pcmag.com/news/twitter-our-read-before-you-retweet-function-actually-works>> last accessed 28 April 2021

¹¹⁶⁰ Anthony Lantian and others, ‘Measuring Belief in Conspiracy Theories: Validation of a French and English Single-Item Scale’ (2016) 29(1) *International Review of Social Psychology* 1, 3

¹¹⁶¹ Mean = 4.81, Standard Deviation = 1.312

¹¹⁶² Part IV, ‘Methodology’ – Research Design; Walter Lippmann, *Public Opinion* (Simon & Schuster 1997) 28, 29

¹¹⁶³ Discussed in Part III, ‘The Social Psychology of Political News and Social Media’

¹¹⁶⁴ *Republicans*, Mean = 5.14, Standard Deviation = 1.219; *Democrats*, Mean = 4.59, Standard Deviation = 1.327

A further follow up item asked participants to visually assess on a sliding percentage scale (0% to 100%) how differently or similarly they saw themselves to their friends and contacts on social media in broad terms of ‘beliefs.’¹¹⁶⁵ Rather than a binary, post hoc determination of whether participants’ social networks reflect an ‘echo chamber,’¹¹⁶⁶ allowing participants to decide for themselves how alike they think they are to their social media peers can provide, at least for my sample, an alternative account of how groups interact and overlap online. The comparison between the groups’ peer similarity reports is shown in **Table C1**. While my Democrat-group participants ranked higher on this item, both groups reported fairly moderate peer similarity evaluations on average.¹¹⁶⁷

Table C1

Comparisons Between Participant Groups by Additional Measures

	Republicans (n = 139)	Democrats (n = 218)		
	<i>Mean Rank</i>	<i>Mean Rank</i>	<i>Z-value</i>	<i>P</i>
Reading Habits	182.95	176.48	-.645	.519
Conspiracism	204.78	160.77	-4.086	<.001
Peer Similarity	158.60	192.00	-2.983	.003

I made a further analysis of these scores (reading habits, conspiracism, and peer similarity) to reveal potential interaction with my participants’ political group identification scores. The Kendall rank correlations for each score are shown in **Table C2**. Average group identification score was positively associated with peer similarity but did not show a significant relationship with conspiracism or reading habits. The association between peer similarity and group identification supports the notion that social identification processes play out through social media. Per the concepts discussed in Part Three,¹¹⁶⁸ my more partisan participants appear to prefer social media interactions with ideologically similar others (such as fellow ingroup members) that would likely reinforce positive group image and group distinctiveness. Among those participants for whom political identity was less central, social media interactions may accordingly be less relevant to their political group identity and thus conceivably welcome exposure to a broader range of beliefs.

¹¹⁶⁵ Mean = 60.25, Standard Deviation = 21.697

¹¹⁶⁶ Discussed in Part I, ‘Social Media News Use’ and Part I, ‘Fake News Studies’

¹¹⁶⁷ *Republicans*, Mean = 56.38, Std. Deviation = 22.221; *Democrats*, Mean = 62.72, Std. Deviation = 21.038

¹¹⁶⁸ Part III, ‘Principles of Social Psychology’

Table C2*Kendall Rank Correlation Table: Additional Measurements*

	Group Identification	Reading Habits	Conspiracism	Peer Similarity
Group Identification	—			
Reading Habits	-.007	—		
Conspiracism	-.009	-.052	—	
Peer Similarity	.221**	-.029	.026	—

Note. n = 357, ** p < .01

Accepting the limitations of these additional scales as discussed above, the correlational analyses do not provide support for the proposition that group identification is related to two of the other factors, reading habits and conspiracism.¹¹⁶⁹ The relationship with peer similarity makes conceptual sense in terms of group identification as an expression of the social identity drives for ingroup positivity and distinctiveness—the latter in particular.

The last additional measurement asked for participants' voting intentions in regard to the then-upcoming 2020 U.S. Presidential election. When asked to indicate their present voting intentions regarding the then-upcoming election (about a month away at the time of the study), participants gave responses that were significantly related to their party-closeness choice, and that relationship was of considerable strength.¹¹⁷⁰ The chi-square analyses performed on these measurements are shown in **Table C3** below. These results support the validity of my two party groupings and fit well with the notion that independent 'leaners' nevertheless tend to vote consistently with their broader political groups.¹¹⁷¹

¹¹⁶⁹ As discussed by Lantian and others (n 1160), conspiracism has generated substantial research interest as an independent factor in its own right

¹¹⁷⁰ Chi-square (Cramer's V = 0.849, p < 0.001)

¹¹⁷¹ Discussed in Part IV, 'Methodology' – Test Procedure: Measuring Social Identification

Table C3*Voting Intention Frequency by Party Grouping*

	Voting Intention					χ^2	<i>Cramer's V</i>
	Donald Trump	Joe Biden	Other	Not Sure	I Would Not Vote		
Republicans (df = 4, n = 139)	113	7	6	5	8	—	—
Democrats (df = 4, n = 218)	7	188	9	5	9	257.421	.849**

Note. ** p < .01

Supporting Analyses

Expanded Correlational Analysis

Table C4

Kendall Rank Correlation Table: Demographic and Test Variables

	Group Identification Score	Sharing Fake Outgroup Stories	Flagging Genuine Ingroup Stories	Age	Annual Gross Income	Sex/Gender	Voter Registration Status
Group Identification Score	—						
Sharing Fake Outgroup Stories	.140**	—					
Flagging Genuine Ingroup Stories	.130**	.458**	—				
Age	-.007	-.087*	-.067	—			
Annual Gross Income	.092*	-.003	-.016	-.045	—		
Sex/Gender	.111*	-.190**	-.148**	.084	-.022	—	
Voter Registration Status	.120**	.021	.071	-.025	.019	-.020	—

Note. * $p < .05$, ** $p < .01$

Weighted Regression

Examining the plots of residual values versus predictor values for my group separated regression analyses of sharing fake outgroup stories and flagging ingroup stories indicated possible non-constant variance: low-end values of average group identification showed markedly less spread than high-end values. Accordingly, I developed weighted least squares regression models by first calculating the absolute values of the original regression analysis residuals and regressed the absolute values against the predictor variables from my original models. I then calculated the weights by taking the reciprocals of the squared values returned by regressing the absolute values against the predictors (*i.e.*, the estimators of the error standard deviations). The group separated weighted regression models for outgroup sharing are shown below in **Table C4** and for ingroup flagging in **Table C5**.

Table C5

Group separated weighted least squares regression analyses predicting willingness to share fake outgroup stories from demographic factors and group identification scores

<i>Republican Participants (n = 139)</i>						
Variable	B	<i>t</i>	sr ²	<i>F</i>	<i>R</i> ²	Δ <i>R</i> ²
Step 1				2.238	.032	—
Age	-.155	-1.809				
Sex/Gender	-.068	-.791				
Step 2				17.884***	.284	.253***
Age	-.154	-2.092*	.023†			
Sex/Gender	-.141	-1.894	.019†			
Group Identification	.508	6.902***	.253			
<i>Democrat Participants (n = 218)</i>						
Variable	B	<i>t</i>	sr ²	<i>F</i>	<i>R</i> ²	Δ <i>R</i> ²
Step 1				10.143***	.086	—
Age	-.075	-1.148				
Sex/Gender	-.277	-4.238***				
Step 2				7.868***	.099	.013
Age	-.067	-1.024	.004†			
Sex/Gender	-.287	-4.395***	.081			
Group Identification	.115	1.766	.013†			

Note. **p* < .05; ***p* < .01; *** *p* < .001; †observed power < .80

The weighted regressions for sharing fake outgroup stories generally confirm the analyses presented in **Table 9** while refining the coefficients and *p*-values obtained. Group identification continues to show marked differences in coefficient value and explanation of variance between the groups.¹¹⁷² Age retains a significant relationship among Republican participants but at a too-low effect size for reliable determination. Among Democrat participants, sex/gender retains a significant relationship with outgroup sharing, while the *p*-value for group identification increases to .079.

¹¹⁷² As discussed in Part IV, 'Findings' – Discussion, why this difference appears is interesting in its own right

Table C6

Group separated weighted least squares regression analyses predicting willingness to flag genuine ingroup stories as fake from demographic factors and group identification scores

<i>Republican Participants (n = 139)</i>						
Variable	B	<i>t</i>	sr ²	<i>F</i>	<i>R</i> ²	Δ <i>R</i> ²
Step 1				9.204***	.119	—
Age	-.227	-2.686**				
Sex/Gender	-.200	-2.360*				
Step 2				23.356***	.342	.222***
Age	-.017	-.208	.000†			
Sex/Gender	-.137	-1.854	.017†			
Group Identification	.528	6.754***	.223			
<i>Democrat Participants (n = 218)</i>						
Variable	B	<i>t</i>	sr ²	<i>F</i>	<i>R</i> ²	Δ <i>R</i> ²
Step 1				35.333***	.247	—
Age	-.400	-6.705***				
Sex/Gender	-.352	-5.894***				
Step 2				47.749***	.401	.154***
Age	-.032	-.436	.001†			
Sex/Gender	-.183	-3.152**	.028			
Group Identification	.549	7.408***	.154			

Note. **p* < .05; ***p* < .01; *** *p* ≤ .001; †observed power < .80

The weighted regressions for flagging genuine ingroup stories likewise complement the analyses shown in **Table 12**. Group identification shows a clearer relationship with outgroup sharing, explaining a substantially greater degree of the variation for both participant groups than in the original analyses. Among Republican participants, age no longer shows a significant relationship (notwithstanding low observed power to begin with), while sex/gender shows a more muted relationship among Democrat participants (though still detectable and significant).

Future Research Supplementals

Intergroup Emotions

In addition to social identity theory, intergroup emotions theory presents an interesting framework for evaluating intergroup behaviour by providing a further account of motivation and action. As a supplemental part of my study, I incorporated two additional question banks using Smith, Seger, and Mackie's twelve emotion-scale items to collect data on group-level emotions from my participants.¹¹⁷³ While not part of my primary investigation, further studies could be constructed around similar data to investigate how particular group emotions may relate to social media decision making. The emotional reactions of experiencers of a public affairs news story on social media in the wild could cover a wide gamut, which made this research area challenging to incorporate in my group identification design without overcomplicating my participants' tasks.

Rather than ranking the participants' experienced intensity of a particular feeling or other methods of measuring emotions, I decided only to test whether participants with higher or lower identification would tend to experience the same or different emotional reactions to the test items.¹¹⁷⁴ As noted in my discussion of social psychology research,¹¹⁷⁵ sense of well-being and mood can be distinctly influenced by group membership, including with regard to political engagement, as can self-esteem and feelings—each of which tend to influence individuals' behaviour. One approach to capturing those factors would be to allow participants to fill in an open text field with a brief expression of what they felt, provided they would be willing to do so. However, that potentially infinite set, along with the necessity to parse or reject ambiguous responses, would also contribute a substantial data processing load. What I was most interested in was whether I would see a commonality of responses among participant groups, rather than looking for any particular reaction. Keeping in mind that this would be strictly a preliminary exploration, I decided that I could best get comparable reactions from my participants by giving them an array of responses to choose from.

Along with the flagging and sharing scales, a six-choice array of emotions was provided along with the instruction to pick the one that best described how the participant felt. This array was adapted from the twelve emotion items used by Smith, Seger, and

¹¹⁷³ Eliot R Smith, Charles R Seger, & Diane M Mackie, 'Can Emotions Be Truly Group Level? Evidence Regarding Four Conceptual Criteria' (2007) 93(3) J Personality and Social Psychology 431, 434

¹¹⁷⁴ Previous studies suggest that those with higher group identification would converge more strongly toward a typical pattern of group emotions; see Charles R Seger, Eliot R Smith, and Diane M Mackie, 'Subtle activation of a social categorization triggers group-level emotions' (2009) 45 J Experimental Psychology 460

¹¹⁷⁵ Part III, 'Principles of Social Psychology' – Social Identity Theory

Mackie.¹¹⁷⁶ Since the news stories all have a negative tone, I removed some of the less relevant choices like “Happy” or “Proud” to avoid presenting the participants with too many options. “Hopeful” and “Satisfied” were retained to facilitate consistency checking and to provide relevant options for participants reacting to stories about their political outgroup. A simplified ‘pick one’ item was used (rather than the seven point scale items in Smith, Seger, and Mackie¹¹⁷⁷) in keeping with the kind of data I wanted to collect.¹¹⁷⁸ These arrays appeared as follows:

Which of the following **best** describes how you felt after reading this story?

- ☐ Hopeful
- ☐ Irritated
- ☐ Satisfied
- ☐ Guilty
- ☐ Disgusted
- ☐ Uneasy

The emotional response block measures showed fair reliability.¹¹⁷⁹ The chi-square analyses¹¹⁸⁰ performed on these measurements are shown in **Table C7** below. The emotional response blocks results showed that for each of the ten Republican stories and ten Democrat stories, participants in each group tended to select certain emotional responses to either category of stories more frequently than others.¹¹⁸¹ To analyse the relationship between group identification and participants’ emotional reactions to the news stories, I looked at whether the responses differed between participants who were in the top half of group identification scores and the bottom half.¹¹⁸²

¹¹⁷⁶ Smith, Seger, and Mackie (n 1173) 434

¹¹⁷⁷ *ibid*

¹¹⁷⁸ A limited selection also partly mimics the handful of reaction ‘emoji’ users can pick on Facebook’s News Feed, although the Facebook selection only provides the ‘Sad’ and ‘Angry’ emoji as negative reactions

¹¹⁷⁹ (Cronbach’s Alpha = 0.801)

¹¹⁸⁰ **Chi-square test:** measures the association between nominal variables (such as my emotional response indicators); a significant Cramer’s V statistic ($p < 0.05$) in this test indicates the variables are substantially associated, with higher values indicating a stronger association between them

¹¹⁸¹ *Republican*, Chi-square (Cramer’s V = 0.235, $p = 0.001$); *Democrat*, Chi-square (Cramer’s V = 0.243, $p < 0.001$)

¹¹⁸² Divided at the median group identification score, 5.50

Table C7*Emotional Reaction Frequency by News Item Category*

	Emotional Reactions						χ^2	<i>Cramer's V</i>
	Hopeful	Irritated	Satisfied	Guilty	Disgusted	Uneasy		
<i>Republican Stories</i>								
All Republicans ^{†a}	8	65	18	0	32	16	—	—
All Democrats ^{†b}	6	92	9	2	87	22	19.784	.235**
High Republicans ^{†c}	8	30	11	0	13	13	—	—
High Democrats ^{†d}	3	50	2	2	49	11	28.762	.387**
Low Republicans ^{††e}	0	35	7	0	19	3	—	—
Low Democrats ^{††f}	3	42	7	0	38	11	6.575	.200
<i>Democrat Stories</i>								
All Republicans ^{†a}	7	43	18	2	61	8	—	—
All Democrats ^{†b}	8	86	10	5	72	37	21.123	.243**
High Republicans ^{†c}	7	20	12	1	31	4	—	—
High Democrats ^{†d}	6	41	7	3	38	22	14.290	.273*
Low Republicans ^{†e}	0	23	6	1	30	4	—	—
Low Democrats ^{†f}	2	45	3	2	34	15	9.237	.237

Note. * $p < .05$, ** $p < .01$, [†] $df = 5$, ^{††} $df = 4$, ^a $n = 139$, ^b $n = 219$, ^c $n = 75$, ^d $n = 117$, ^e $n = 64$, ^f $n = 101$

Among participants with lower group identification scores, I found no significant association between the emotional responses picked by members of each group regarding Republican or Democrat news stories.¹¹⁸³ In other words, among less strongly partisan Republicans or Democrats, news about one party or the other did not tend to correspond

¹¹⁸³ *Republican*, Chi-square (Cramer's V = 0.200, $p = 0.160$); *Democrat*, Chi-square (Cramer's V = 0.237, $p = 0.100$)

with distinctly different emotional responses. But among participants in the top half of group identification scores, I found a significant difference between the emotional responses each party group picked when presented with news about their own party or the other side—the responses picked by highly identified participants showed substantial association.¹¹⁸⁴

Ideological Asymmetry

A future study could also focus more closely on potential ideological asymmetries between participant groups as covariant or independent explanations for observed decision making behaviours. In my study, another supplemental question presented participants with a bank of twelve political issues, each accompanied by seven point-scale items anchored at point one by “Not at all important” and at point seven by “Very important.” I took these measurements to provide additional support for the notion that the participants’ political groups meaningfully reflect real *political* differences—distinct priorities of the groups regarding matters of public concern or government action. Four of the issues were taken directly from the Republican National Committee’s statement of principles,¹¹⁸⁵ four were taken from the Democratic National Committee’s platform,¹¹⁸⁶ and four were issues about public discourse common to both parties, included specifically for their relevance to this study.¹¹⁸⁷ Reliability across all twelve issue scales was modest,¹¹⁸⁸ but higher among the respective partisan issues¹¹⁸⁹ than the four public discourse ones.¹¹⁹⁰

I compared my participants’ issue importance ratings using the Mann-Whitney test and found a significant difference between my Republican and Democrat participants on each of the partisan issues as expected, indicating that my participant groups are properly sorted by party affiliation. Those findings are presented in **Table C8** below. The clear divide in partisan issue importance suggests my experiment design could be expanded by tying story prompt content to party-specific issues in order to engage related political psychologies. While I expected to find a significant difference between the groups on the partisan issues, I was surprised to find a significant difference between the groups on two of

¹¹⁸⁴ *Republican*, Chi-square (Cramers’ V = 0.387, $p < 0.001$); *Democrat*, Chi-square (Cramer’s V = 0.273, $p = 0.014$)

¹¹⁸⁵ Republican National Committee, ‘Principles for American Renewal’ (2020) <<https://www.gop.com/principles-for-american-renewal>> last accessed 4 May 2021 (namely, “Balanced budget amendment,” “School choice,” “Securing the border,” and “Religious liberty”)

¹¹⁸⁶ Democratic National Committee, ‘The 2020 Democratic Platform’ (2020) <<https://democrats.org/where-we-stand/party-platform/>> last accessed 4 May 2021 (specifically, “Anti-discrimination,” “The Affordable Care Act,” “Gun control,” and “Fighting climate change”)

¹¹⁸⁷ Namely, “Relying on the facts,” “Hearing both sides,” “Freedom of speech,” and “Freedom of the press”
¹¹⁸⁸ (Cronbach’s Alpha = 0.685)

¹¹⁸⁹ *Republican* (Cronbach’s Alpha = 0.703); *Democrat* (Cronbach’s Alpha = 0.834)

¹¹⁹⁰ *Public Discourse* (Cronbach’s Alpha = 0.666)

the (assumedly non-partisan) public discourse issues (which both groups nevertheless rated as highly important on average). The groups broke in opposite directions on the issues of freedom of the press¹¹⁹¹ and hearing both sides,¹¹⁹² perhaps speaking to different ideas about how the press should function in relation to political debate, as discussed in Part Five.¹¹⁹³ Specific ideological content regarding these discourse issues could also inform comparisons with other motivational factors explored under political psychology frameworks.

Table C8

Comparisons Between Participant Groups by Political Issues

	Political Group		Z-value	p
	Republicans (n = 139)	Democrats (n = 217)		
	Mean Rank	Mean Rank		
Republican Issues				
Balanced Budget	194.03	167.81	-2.381	.017
School Choice	203.90	161.53	-3.846	<.001
Border Security	248.94	132.89	-10.518	<.001
Religious Liberty	221.62	150.26	-6.608	<.001
Democrat Issues				
Anti-discrimination	124.68	212.98	-8.299	<.001
Affordable Care Act	117.34	217.68	-9.296	<.001
Gun Control	132.67	207.86	-6.903	<.001
Climate Change	109.79	222.51	-10.536	<.001
Public Discourse Issues				
Relying on facts	171.70	182.85	-1.234	.217
Freedom of speech	187.94	172.46	-1.535	.125
Freedom of the press	165.06	187.11	-2.122	.034
Hearing both sides	192.23	169.71	-2.174	.030

Note. For the *Republican Issues* ranks, Republican participant group n = 138

¹¹⁹¹ Mann-Whitney ($Z = -2.122$, $p = 0.034$)

¹¹⁹² Mann-Whitney ($Z = -2.174$, $p = 0.030$)

¹¹⁹³ Part V, 'De-mystifying Social Media' – Media Systems Analysis, – Social Media and the Public Sphere

Further research incorporating ideological content concerning public discourse may be particularly useful in guiding public efforts to promote a healthy news ecosystem. For example, fact checking, as discussed elsewhere in this Thesis, conceivably relates to ‘relying on facts’ as an important matter of party ‘doctrine’ among Republicans and Democrats alike, going by the data I collected. When I performed a Kendall Rank analysis including ‘relying on facts’ alongside my primary test variables, I found several interesting associations, as shown in **Table C9** below. Among my Republican participants, relying on facts was not significantly associated with average sharing scores for fake outgroup stories, but among my Democrat participants, the association was significant and negative. Among both groups, however, relying on facts was significantly negatively associated with average flagging scores for genuine ingroup stories.

Table C9

Kendall Rank Correlation Table: Additional Public Discourse Issue

	Group Identification	Sharing Fake Outgroup Stories	Flagging Genuine Ingroup Stories	Relying on Facts
<i>Republican Participants</i>				
Group Identification	—			
Sharing Fake Outgroup Stories	.244**	—		
Flagging Genuine Ingroup Stories	.151*	.537**	—	
Relying on Facts	-.048	-.079	-.177*	—
<i>Democrat Participants</i>				
Group Identification	—			
Sharing Fake Outgroup Stories	.074	—		
Flagging Genuine Ingroup Stories	.124*	.397**	—	
Relying on Facts	.112*	-.147**	-.241**	—

Note. * $p < .05$, ** $p < .01$

I next performed group separated regression analyses to see whether relying on facts similarly explained some of the variance in sharing fake outgroup stories, which was difficult to characterize in terms of relation to group identification. In Step 1 of these regressions, I included age (centered) and sex/gender, as before, as well as group identification score. In

Step 2, I added participants' scores for the importance of relying on facts (centered). The results of these regressions are shown below in **Table C10**.

Table C10

Group separated regression analyses predicting willingness to share fake outgroup stories from demographic factors, group identification scores, and importance of relying on facts

<i>Republican Participants (n = 139)</i>						
Variable	B	<i>t</i>	sr ²	<i>F</i>	<i>R</i> ²	Δ <i>R</i> ²
Step 1				11.200***	.199	—
Age	-.189	-2.437*				
Sex/Gender	-.152	-1.949				
Group Identification	.397	5.108***				
Step 2				8.567***	.204	.004
Age	-.203	-2.561*	.047			
Sex/Gender	-.153	-1.959	.027†			
Group Identification	.400	5.143***	.165			
Relying on Facts	.068	.858	.005†			
<i>Democrat Participants (n = 218)</i>						
Variable	B	<i>t</i>	sr ²	<i>F</i>	<i>R</i> ²	Δ <i>R</i> ²
Step 1				9.003***	.113	—
Age	-.065	-1.001				
Sex/Gender	-.310	-4.754***				
Group Identification	.124	1.903				
Step 2				6.904***	.115	.003
Age	-.060	-.913	.004†			
Sex/Gender	-.302	-4.556***	.089			
Group Identification	.128	1.958	.018†			
Relying on Facts	-.053	-.807	.003†			

Note. **p* < .05; ***p* < .01; *** *p* < .001; †observed power < .80

The results suggest that 'relying on facts' does not significantly explain much difference in sharing fake outgroup stories among Republicans *or* Democrats, despite the latter group's significant negative association between relying on facts with the outgroup

sharing measure as well as its significant positive association between relying on facts and group identification (suggesting its ‘doctrinal’ relevance to Democrats). While this analysis is strictly preliminary, further research in this direction could indicate whether fact checking schemes, as a ‘solution’ to reduce fake news sharing, appropriately engage relevant social or political psychologies—or not.

Sex/Gender Differences

An unexpected finding of my regression analyses was the apparent gender bias in sharing fake outgroup stories and flagging genuine ingroup stories. In my Republican group, participants were 52.5% males and 47.5% females; while in my Democrat group, participants were 47.2% males, 51.8% females, and 0.9% indicating other. The difference between the Republicans and Democrats was not statistically significant, suggesting that the apparent effect was not due to an imbalance in how participants of different genders were distributed between my groups.¹¹⁹⁴ The negative slope of the coefficients in both cases indicates female and non-binary participants tended to score lower on the test measurements. In my initial group separated model for outgroup sharing,¹¹⁹⁵ the sex/gender relationship was significant among Democrat participants and nearly significant among Republican participants at $p = .053$, keeping in mind the low effect size among Republicans. In the initial group separated model for ingroup flagging,¹¹⁹⁶ the sex/gender relationship with the dependent variable was significant for both groups, though again of small effect size relative to the size of my Republican sample. In my weighted regression models, the sex/gender relationship to either variable was not significant among Republicans.¹¹⁹⁷

To better understand the sex/gender relationships, I built separate regression models for each sex/gender response—simple regressions of age (centered), party affiliation, and average group identification score (centered). Party affiliation was coded as 1 for Republicans and 2 for Democrats, so the direction of the coefficients, negative or positive, indicates the relationship with either party (respectively) and the sharing/flagging variable. The results of these sex/gender separated regression analyses¹¹⁹⁸ are shown in **Table C11** for sharing fake outgroup stories and in **Table C12** for flagging genuine ingroup stories.

With regard to sharing fake outgroup stories, the regression analyses showed that party affiliation was significantly negatively related to the test measure among female

¹¹⁹⁴ Mann-Whitney ($Z = -1.047$, $p = .295$)

¹¹⁹⁵ See **Table 9**

¹¹⁹⁶ See **Table 12**

¹¹⁹⁷ See **Table C5** and **Table C6**

¹¹⁹⁸ In both instances, only male and female participant responses have been analyzed, as the two participants who did not identify as male or female both indicated the same party affiliation

participants, suggesting a distinct difference between female Republicans and female Democrats, albeit at an effect size slightly below reliable discernment at the size of my sample.¹¹⁹⁹ Party affiliation did not show a significant relationship to the test variable among male participants, suggesting little difference between party groups among men. Group identification was positively related to outgroup sharing for both male and female participants, but the coefficient and effect size was larger among the male group.

Table C11

Sex/gender separated regression analyses predicting willingness to share fake outgroup stories from age, party affiliation, and group identification scores

<i>Male Participants (n = 176)</i>					
Variable	B	t	sr ²	F	R ²
				6.517***	.102
Age	-.127	-1.742	.015†		
Party Affiliation	-.032	-.441	.001†		
Group Identification	.298	4.117***	.089		
<i>Female Participants (n = 179)</i>					
Variable	B	t	sr ²	F	R ²
				5.246**	.083
Age	-.094	-1.291	.008†		
Party Affiliation	-.196	-2.686**	.038†		
Group Identification	.200	2.746**	.040†		

Note. *p < .05; **p < .01; *** p < .001; †observed power < .80

With regard to flagging genuine ingroup stories as fake, the regression analyses showed again that party affiliation (coded 1 for Republicans, 2 for Democrats) was significantly negatively related to the test measure among female participants but not among male participants, though at an even smaller effect size than in the outgroup sharing condition. Nevertheless, party affiliation appears to matter in whether women share fake news about the other side or flag genuine stories about their own side as fake. Group identification showed a stronger significant positive relationship to ingroup flagging among male participants compared with female participants, among whom the effect size was much smaller.

¹¹⁹⁹ Cohen's f^2 = .040, observed power = .75

Table C12

Sex/gender separated regression analyses predicting willingness to flag genuine ingroup stories from age, party affiliation, and group identification scores

<i>Male Participants (n = 176)</i>					
Variable	B	<i>t</i>	sr ²	<i>F</i>	<i>R</i> ²
				9.214***	.138
Age	-.190	-2.660**	.035†		
Party Affiliation	-.088	-1.238	.008†		
Group Identification	.323	4.556***	.104		
<i>Female Participants (n = 179)</i>					
Variable	B	<i>t</i>	sr ²	<i>F</i>	<i>R</i> ²
				2.970*	.048
Age	-.039	-.530	.002†		
Party Affiliation	-.153	-2.062*	.023†		
Group Identification	.164	2.201*	.026†		

Note. **p* < .05; ***p* < .01; *** *p* < .001; †observed power < .80

Appendix D

20th Century American Jurisprudence

The aim of this Appendix is to provide additional background on key developments of American speech and communications law that took place in the 20th Century. British and European readers may find this exploration helpful in laying bare the roots of present U.S. legal thinking about the appropriateness of government intervention re media use. In the first section, I cover a series of cases that arose following the 1917 Espionage Act and its amendments. These cases were integral to the later development of the *Brandenburg* test for the constitutionality of content-based speech and press restrictions. In the second section, I discuss a series of cases relating to the 1927 Radio Act and the burgeoning role of federal regulatory agencies. This period features the development and articulation of the American ‘public interest’ standard for broadcast regulation. In the final section, I evaluate the progression of American broadcast regulation in light of subsequent television industry development and a pair of cases concerning the 1992 Cable Act. Each of these phases of jurisprudence reflects a particular perspective on where and how decisions about communications content should be made.

Holdovers of 19th Century Liberalism

In this section, I discuss some of the Espionage Act cases that illustrate the Supreme Court’s thinking at the time. During this phase, the Court was attempting to reconcile 19th Century liberal principles toward speech with 20th Century exigencies—World War I, the Russian Revolution, and an increasingly ambitious national government.

The first of these cases involved a socialist political organizer named Schenck whose party had published and mailed out leaflets urging resistance to conscription into the American Expeditionary Force.¹²⁰⁰ Schenck challenged his conviction under the Espionage Act on First Amendment grounds, but the Supreme Court unanimously held against him, declaring that the context of the speech act determines whether an utterance or publication is of tolerable character—if the words used are “used in such circumstances and are of such a nature as to create a clear and present danger that they will bring about the substantive evils that Congress has a right to prevent.”¹²⁰¹ The decision, written by Justice Oliver Wendell Holmes, Jr., goes on to describe “circumstance” and “nature” in terms of proximity and

¹²⁰⁰ *Schenck v. U.S.* [1919] 249 U.S. 47 (Schenck’s case was combined with that of Baer, a fellow member of the Socialist party executive committee with Schenck who faced the same charges)

¹²⁰¹ *Schenck*, 249 U.S. at 52; consider further that the ‘right’ of the U.S. Congress to do anything is only that which is expressly provided to it by the Constitution

degree (terms which themselves would be refined in subsequent cases)—in this case, the actuality of the Congressional effort to raise an army and the admitted likelihood of Schenck’s leaflet to have some effect against the effort’s success. Holmes provides the sensible example of “falsely shouting fire in a theatre and causing a panic” as speech which would clearly not be protected but recognizes that outside of wartime Schenck’s statements might otherwise be permissible.

In 1919’s *Abrams v. United States*, however, Justice Holmes would find himself writing the dissent to the Court’s opinion.¹²⁰² *Abrams* involved five defendants charged under the disloyal abuse provision of the amended Espionage Act, language intended to bring the government into disrepute, encouragement of resistance to the United States, and the discouragement of producing war materiel.¹²⁰³¹²⁰⁴ Relying on *Schenck*, Justice Clarke wrote for the majority, reasoning that “Men must be held to have intended, and to be accountable for, the effects which their acts were likely to produce.”¹²⁰⁵ Justice Holmes, joined by Justice Brandeis in the dissent, pointed out that the group’s leaflets denounced German militarism, made no attack on the *form* of American government, and did not meet the Espionage Act’s required showing of intent thereby to “cripple or hinder” the U.S. in prosecuting the war.¹²⁰⁶

Holmes argued for a strict reading of “actual intent” in the statute to avoid criminalizing instances where a person might advocate ending wasteful spending on a particular military vehicle or cannon, and such advocacy could result in the curtailment of such production, without any intent to hinder the war effort (rather, in the example, to improve it). Holmes explains that such a reading of intent is further necessary to make sense of an *attempted* violation of the Act, seeing as some general intent (such as defending the Russian Revolution) could be accomplished “without bringing about the evils” Congress sought to check (i.e., hindrance of the war effort)—and thus, despite possibly producing some indirect or undesired effect, not posing the kind of present danger that would justify

¹²⁰² *Abrams v. U.S.* [1919] 250 U.S. 616 (Holmes’ dissent begins on 624)

¹²⁰³ *ibid* 617

¹²⁰⁴ *ibid* 619-622 (Abrams and his co-defendants were Russian-born but had all lived in the U.S. for five to ten years, and while several of them were self-avowed “revolutionists,” the gist of their offending leaflets was to point out President Wilson’s hypocrisy in going to war and suggest “cowardly” capitalist motivations in undermining the nascent revolution in Russia, declare that all workers had a common enemy in capitalism, and urge immigrant workers in particular to “throw away all confidence” in the government that had asked for their help in prosecuting the war and go on strike)

¹²⁰⁵ *ibid* 621 (the defendants’ aim to aid the cause of the Russian Revolution could be of no avail due to the fact that the effect of their appeal would be to undermine the U.S. government’s war in Europe, especially if workers in an ammunition factory were to go on strike)

¹²⁰⁶ *ibid* 625-26

interference with the right to free speech.¹²⁰⁷ Similar discrimination would be necessary in the interpretation of statutes concerning social media communications to avoid drawing in hosts of users casually commenting on stories and discussing political events.

Holmes' dissent concludes with a passage that, in discussions of social media, has been frequently quoted by advocates and opponents of online free speech, who point to the notion of a social media 'marketplace of ideas' as a wellspring to be protected from burdensome government interference in free thought, or alternatively as a failure to be corrected by state intervention to produce better ideas. Both perspectives mistake the business-oriented vernacular of Holmes' day for the point he was making,¹²⁰⁸ which is otherwise quite clear from the context (emphasis mine):

"Persecution for the expression of opinions seems to me perfectly logical. If you have no doubt of your premises or your power, and want a certain result with all your heart, you naturally express your wishes in law, and sweep away all opposition. To allow opposition by speech seems to indicate that you think the speech impotent . . . , or that you do not care wholeheartedly for the result, or that you doubt either your power or your premises. But when men have realized that time has upset many fighting faiths, they may come to believe *even more than they believe the very foundations of their own conduct* that the ultimate good desired is better reached by free trade in ideas -- that the best test of truth is the power of the thought to get itself accepted in the competition of the market, and that truth is the only ground upon which their wishes safely can be carried out.

In its full context, Holmes is stating the moral basis upon which power is legitimated in democratic societies—legitimate power is not derived from conviction but from reason. Reason is demonstrated when premises are tested for truth. Holmes continues (again, emphasis mine):

That, at any rate, is the theory of our Constitution. It is an experiment, as all life is an experiment. Every year, if not every day, we have to wager our salvation upon some prophecy based upon imperfect knowledge. While that

¹²⁰⁷ *Abrams*, 250 U.S. at 627-29 (Holmes also intimates that the severe 20-year prison sentence facing the defendants seemed to have more to do with their creed—of which Holmes patently disapproved—than the gravity of their offence.)

¹²⁰⁸ Holmes is not making a claim about the efficiency of markets, an economics theory that would not be developed for several more decades; see, e.g., Friedrich A Hayek, 'The Use of Knowledge in Society' (September 1945) 35(4) *American Economic Review* 519; a judge writing in today's scientistic parlance might call social media the 'laboratory of ideas' and suggest that the best test for truth is to withstand peer review

experiment is part of our system, I think that we should be *eternally vigilant* against attempts to check the expression of opinions that we loathe and believe to be fraught with death, unless they so imminently threaten immediate interference with the lawful and pressing purposes of the law that an immediate check is required to save the country. I wholly disagree with the argument of the Government that the First Amendment left the common law as to seditious libel in force. History seems to me against the notion. I had conceived that the United States, through many years, had shown its repentance for the Sedition Act of 1798, by repaying fines that it imposed. Only the emergency that makes it *immediately dangerous to leave the correction of evil counsels to time* warrants making any exception to the sweeping command, 'Congress shall make no law . . . abridging the freedom of speech.' Of course, I am speaking only of expressions of opinion and exhortations, which were all that were uttered here, but I regret that I cannot put into more impressive words my belief that, in their conviction upon this indictment, the defendants were deprived of their rights under the Constitution of the United States."¹²⁰⁹

Holmes' writing here is worth including at length because of how comprehensively he captures the circumstances that arise when authority faces opposing narratives. Banning contrary opinions as 'fake news' has a seductive logic and satisfies a range of different interests, while submitting one's own accounts of political events to the constraint of free inquiry risks much and demands both confidence and humility. As noted by Ross, it was probably "no accident" that the Supreme Court saw case after case involving laws directed against left-wing causes in the following decade.¹²¹⁰ Moreover, Holmes is counselling against unprincipled laws like the Espionage Act and the Sedition Act before it¹²¹¹ because unprincipled laws tend to evoke distrust and disloyalty towards the governments that enact them.

Holmes would end up in the dissent to further cases,¹²¹² including *Gitlow v. New York* in 1925, where he would again push for a narrower treatment of incitement offences.¹²¹³ But by *Gitlow*, the majority of the Court had acknowledged (at least in *dicta*) that protection of the freedoms of the First Amendment applied also to the actions of the several states through

¹²⁰⁹ *Abrams*, 250 U.S. at 630-631

¹²¹⁰ William G Ross, *The Chief Justiceship of Charles Evans Hughes, 1930-1941* (U of S Carolina P 2007) 192

¹²¹¹ Discussed in Part V, 'De-mystifying Social Media' – Advancements in Law

¹²¹² *e.g.*, *Schaefer v. United States* [1920] 251 U.S. 466; *Pierce v. United States* [1920] 252 U.S. 239

¹²¹³ *Gitlow v. New York* [1925] 268 U.S. 652, 672-73 ("Every idea is an incitement. . . . Eloquence may set fire to reason. But whatever may be thought of the redundant discourse before us, it had no chance of starting a present conflagration.")

the Due Process Clause of the Fourteenth Amendment.¹²¹⁴ In *Whitney v. California*,¹²¹⁵ the Court found no due process violation,¹²¹⁶ but the circumstances of the case gave Justice Brandeis the chance to further develop the standard by which courts could evaluate whether infringement of the rights of free speech, assembly, and association might be warranted.¹²¹⁷ Building on the necessity of a “clear and present danger” to justify infringing fundamental rights as espoused by Holmes in *Schenck*,¹²¹⁸ Brandeis urged the Court to fix as its standard an *imminent* incidence of purported danger (“so imminent that it may befall before there is opportunity for full discussion” or “an emergency”) and that the imminent harm must also be “relatively serious” (carrying the “probability of serious injury to the State”).¹²¹⁹ In 1969, the Court would hear the case of *Brandenburg v. Ohio*,¹²²⁰ which featured a nearly identically worded criminal syndicalism statute as California’s in *Whitney* decades earlier. By that time, the Court had come around to Brandeis’ proposed tighter standard, that “the constitutional guarantees of free speech and free press do not permit a State to forbid or proscribe advocacy of the use of force or of law violation except where such advocacy is directed to inciting or producing imminent lawless action and is likely to incite or produce such action,”¹²²¹ and struck down the Ohio law.¹²²²

Transitions to Industry Regulation

The cases discussed in this section pertain to the development of law around the newly created regulatory agencies of the 1920s and 30s that governed radio and later television broadcasting in the U.S.—the Federal Radio Commission and its successor, the Federal Communications Commission. During this phase, new thinking came to justify a different level of government involvement with newly organized communications industries.

Under the Radio Act of 1912,¹²²³ radio broadcasting in the U.S. had grown very quickly, accelerating after the end of World War I. News reading was “part of the earliest phase of radio experimentation,” including a US-election returns broadcast in 1916, while “the Harding—Cox election of 1920 is said to mark the start of regular news broadcasting in

¹²¹⁴ *Gitlow*, 268 U.S. at 666

¹²¹⁵ *Whitney v. California* [1927] 274 U.S. 357 (Whitney was a member and organizer of the Communist Labor Party of California; Whitney challenged, on grounds of the Due Process Clause, her conviction under California’s law criminalizing organization of or membership in any group that advocated, taught, or assisted in the commission of crime, sabotage, or unlawful acts or terrorism as a means to effect political or industrial change)

¹²¹⁶ *Whitney*, 274 U.S. at 371

¹²¹⁷ *ibid* 372-380

¹²¹⁸ *Schenck*, 249 U.S. at 52

¹²¹⁹ *Whitney*, 274 U.S. at 374, 377

¹²²⁰ *Brandenburg v. Ohio* [1969] 395 U.S. 444

¹²²¹ *ibid* 447

¹²²² *ibid* 449 (overturning *Whitney*)

¹²²³ Radio Act of 1912, 37 STAT. 302

the US.”¹²²⁴ By 1922, over five hundred stations were operating across the U.S.¹²²⁵ Newspaper industry concerns over competition would lead the American Newspaper Publishers Association to attempt to restrict access by radio stations to its wire services, which were central to national and international news reporting.¹²²⁶ This period of radio development was decades later described by the Supreme Court as having been “left entirely to the private sector” and having resulted in “chaos,”¹²²⁷ citing the experiences of then-Secretary of Commerce Herbert Hoover.¹²²⁸

Hoover had lost two key legal challenges to his authority as radio regulator in 1923 and 1926. In the first, *Hoover, Sec’y Comm. v. Intercity Radio Co., Inc.*,¹²²⁹ Intercity Radio had obtained in early 1920 a license for its wireless telegraphy station in New York City. When its license expired in late 1921, Secretary Hoover refused to renew it, claiming that issuances were at his discretion.¹²³⁰ Hoover had previously attempted (unsuccessfully) to revoke the license of Intercity at the behest of the New York Times and the Navy.¹²³¹ The Court of Appeals for the District of Columbia ordered Hoover to renew the license, explaining that the language of the 1912 Act and the intent of the Congress made clear the Secretary had no discretion to refuse license applications.¹²³² Hoover would subsequently declare that central regulation of radio had become “a public necessity,”¹²³³ a sentiment shared by representatives from companies like Westinghouse who wanted the federal government to impose an arbitrary limit on the total number of broadcasting stations.¹²³⁴ Hoover would suffer a second defeat in *United States v. Zenith Radio Corporation*,¹²³⁵ where the government attempted to enforce against Zenith the terms of its 1925-issued license, which included a restriction on Zenith’s broadcasting times that expressly privileged its competitor General Electric.¹²³⁶ The Court held that the 1912 statute could not be construed to have given the Secretary the arbitrary power to prescribe additional regulations beyond those provided for

¹²²⁴ Brian Winston, *Media technology and society: a history* (Routledge 1998) 85

¹²²⁵ Thomas Hazlett, *The Political Spectrum: The Tumultuous Liberation of Wireless Technology, from Herbert Hoover to the Smartphone* (Yale UP 2017) 45

¹²²⁶ Winston (n 1224) 86

¹²²⁷ *Red Lion Broadcasting Co., Inc. v. FCC* [1969] 395 U.S. 367, 375

¹²²⁸ *ibid* fn. 4

¹²²⁹ *Hoover, Sec’y Comm. v. Intercity Radio Co., Inc.* [1923] 52 App.D.C. 339

¹²³⁰ *ibid* 339

¹²³¹ Marvin R Bensman, ‘Intercity Radio Company Case’ in Donald G Godfrey & Frederic A Leigh (eds.) *Historical Dictionary of American Radio* (Greenwood Press 1998) 212-213

¹²³² *Intercity*, 52 App.D.C. 339

¹²³³ See, e.g., Harvard Law Review Ass’n, ‘The Freedom of Radio Speech,’ (1933) 46(6) Harvard L Rev 987-993

¹²³⁴ Hazlett (n 1225) 45

¹²³⁵ *U.S. v. Zenith Radio Corporation* [1926] 12 F.2d 614

¹²³⁶ *ibid* 616-17

in the Act.¹²³⁷ Hoover would at that point claim he could “no longer police the airwaves” and essentially refused to do so until Congress passed a new Radio Act.¹²³⁸

The regulatory scheme adopted under the 1927 Radio Act¹²³⁹ denied the new Federal Radio Commission any direct power of censorship over radio communications, or of regulation that would interfere with free speech, but instead conditioned the license and relicense of broadcasters on the Commission’s determination of how best to serve the “public interest, convenience or necessity.”¹²⁴⁰ “Public interest” along with “limited spectrum” would become the twin pillars upon which broadcast regulation would stand for decades to follow.¹²⁴¹ Winston points out however that university and educational broadcasters were already operating with missions of civic responsibility well before 1927.¹²⁴² Alabama’s electricity company, for example, operated the State’s first licensed radio station from 1922 with a public service orientation (their call sign “WSY” yielding the slogan “We Serve You”), broadcasting weather and farm reports, interviews, music programs, community bulletins, and church services.¹²⁴³ And as noted by Branscomb, Hoover’s declaration that the physically limited radio frequency spectrum should be publicly managed did not rest on any persuasive legal reasoning: “During the early days of radio interference, the courts could have coped with sorting out property rights to radio frequencies, using the same tools they have used in sorting out property rights in land and other diverse interests. Thus the difference between the print and broadcast media is primarily a historical accident.”¹²⁴⁴ The contrast in American treatment between print and radio content which resulted from Hoover’s declaration would be thrown into sharp relief by two cases in the early 1930s.

*Near v. Minnesota*¹²⁴⁵ involved a state law which made the operation of a “malicious, scandalous and defamatory newspaper” a public nuisance.¹²⁴⁶ Under the statute, a permanent injunction had been obtained against Near’s *Saturday Press* newspaper, which had run a series of stories in late 1927 alleging that various public officials in Minneapolis were in league with local bootleggers and racketeers.¹²⁴⁷ Near denied the stories were malicious, scandalous or defamatory and argued the statute violated his substantive due

¹²³⁷ *Zenith*, 12 F.2d at 617-18

¹²³⁸ Hazlett (n 1225) 41-43

¹²³⁹ Radio Act of 1927, 44 STAT. 1162

¹²⁴⁰ ‘The Freedom of Radio Speech’ (n 1180) 987

¹²⁴¹ As confirmed in *Red Lion*, 395 U.S. at 377

¹²⁴² Winston (n 1224) 81

¹²⁴³ Alabama Historical Radio Society, ‘WSY – First Radio Station in Alabama’ (2020)

<<https://alhrs.org/wsy/>> accessed 6 March 2022

¹²⁴⁴ Anne Branscomb, ‘Citizen Access to the Media’ in William L Rivers & Michael J Nyhan (eds), *Aspen Notebook on Government and the Media* (Praeger London 1973) 120, 125

¹²⁴⁵ *Near v. Minnesota* [1931] 283 U.S. 697

¹²⁴⁶ *ibid* 702 (citing Mason’s Minnesota Statutes, 1927, 10123-1 to 10123-3)

¹²⁴⁷ *ibid* 704-05

process rights.¹²⁴⁸ Chief Justice Charles Evans Hughes, writing the Court's opinion reversing the judgment against *Near*, made clear that the freedom of the press (as of speech) was incorporated in the fundamental liberties protected by the due process clause of the Fourteenth Amendment.¹²⁴⁹ Hughes explained that Minnesota law was "unusual" in its approach at suppressing a newspaper as a public nuisance,¹²⁵⁰ but that the object of the statute was clear from its application: to allow public officials to suppress scandals they could not prosecute as criminal libel.¹²⁵¹ Hughes reasoned further that the law essentially put newspapers on notice that they would be punished for publishing charges of official misconduct, an "effective censorship" inconsistent with the legal tradition holding against prior restraint of the press.¹²⁵²

In the following year, the Court of Appeals of the District of Columbia would hear *Trinity Methodist Church, South v. Federal Radio Commission*,¹²⁵³ which involved the Federal Radio Commission's decision to deny the Church's application to renew its radio station license under the 1927 Radio Act. While the initial examiner had recommended renewal of the license, exceptions were filed,¹²⁵⁴ resulting in the unfavourable re-review of the licensure.¹²⁵⁵ Trinity Church appealed the decision on the grounds, *inter alia*, that the refusal violated the First Amendment guaranty of free speech.¹²⁵⁶ Dr. Bob Shuler, the minister in charge of Trinity Church, was known by his estimated 600,000 listeners as "Fighting Bob" Shuler for his crusades against public corruption in Los Angeles over three years of broadcasting.¹²⁵⁷ Voluminous testimony was adduced before the Commission that Shuler had made various charges in his broadcasts against the Los Angeles courts, bar association, and board of health, as well as having made allegations of bootlegging and gambling.¹²⁵⁸ After a quick gloss of First Amendment tradition (and even a citation to *Near v. Minnesota*), the D.C. Court of Appeals provides the following rejoinder: "But this does not mean that the government, through agencies established by Congress, may not refuse a renewal of license

¹²⁴⁸ *Near*, 283 U.S. at 705

¹²⁴⁹ *ibid* 707

¹²⁵⁰ *ibid* 707

¹²⁵¹ *ibid* 711

¹²⁵² *ibid* 712-15

¹²⁵³ *Trinity Methodist Church, South v. Federal Radio Comm'n* [1932] 62 F.2d 850

¹²⁵⁴ 71 PUB.L. 494, 46 STAT. 844 (a July 1930 amendment of the Radio Act of 1927 allowed for the appeal of decisions of the Commission by licensees *and* by "any other person, firm, or corporation aggrieved or whose interests are adversely affected by any decision of the Commission.")

¹²⁵⁵ *Trinity*, 62 F.2d at 851

¹²⁵⁶ *ibid*

¹²⁵⁷ Charley Orbison, "'Fighting Bob' Shuler: Early Radio Crusader" (1977) 21(4) J Broadcasting 459, 462-63

¹²⁵⁸ *Trinity*, 62 F.2d at 851, 852; Orbison (n 1257) 464, 465 (One of Shuler's campaigns concerned the failed 1928 prosecution of the Julian Petroleum Corporation Ponzi scheme, where Shuler publicly pressured the district attorney and city prosecutor to resign. The subsequently *former* city prosecutor served as part of intervener George Lyon's legal counsel in the appeal of the FRC's decision to renew Shuler's license. Lyon, who had no previous connection to Shuler, had intervened under the amended Radio Act "as a taxpayer and citizen" who was simply "upset with what he felt was Shuler's unfair treatment of his friends.")

to one who has abused it to broadcast defamatory and untrue matter. In that case, there is not a denial of the freedom of speech, but merely the application of the regulatory power of Congress in a field within the scope of its legislative authority.”¹²⁵⁹ In other words, Hoover had ultimately prevailed after his defeats in *Intercity* and *Zenith*.

The decision leaned heavily on the Congressional power to regulate interstate commerce,¹²⁶⁰ and set out an incredibly broad theory of the reach of that power as “no more restricted by the First Amendment than are the police powers of the States under the Fourteenth Amendment.”¹²⁶¹ The cases cited in support of this claim¹²⁶² do not directly stand for the propositions of which the D.C. Court of Appeals has offered them: *In re Kemmler*¹²⁶³ and *Hamilton v. Kentucky*¹²⁶⁴ are both substantive due process cases, with *Kemmler* affirming that the Fourteenth Amendment placed the same “base” underneath the rights to life, liberty, and property before the state governments as the Fifth Amendment placed before the federal government, without any further infringement on state government police power¹²⁶⁵; and with *Hamilton* affirming that the federal government lacked the powers reserved to the states under the Tenth Amendment (such as the police power)¹²⁶⁶ with the caveat that the exercise of the Congress’s valid powers might tend to accomplish a similar purpose “subject to applicable constitutional limitations.”¹²⁶⁷

But despite the obvious applicable constitutional limitation of “Congress shall make no law . . . abridging the freedom of speech, or of the press,”¹²⁶⁸ the D.C. Court of Appeals in *Trinity* suggests that Congress, through the Radio Act of 1927, had acted at the height of its powers, and that the standard of review was no more than a showing of “reasonable exercise of governmental control for the public good.”¹²⁶⁹ If, however, the Constitutional limitation of power among the States is no greater than it is for the federal government, then the principles applied in *Near* against Minnesota’s public nuisance law should have equally condemned any operation of the Radio Act to effect a prior restraint on Shuler’s continued broadcasting.¹²⁷⁰ The appropriate remedy for Shuler’s supposedly defamatory broadcasts

¹²⁵⁹ *Trinity*, 62 F.2d at 851

¹²⁶⁰ U.S. CONST., art. 1, s. 8

¹²⁶¹ *Trinity*, 62 F.2d at 852

¹²⁶² *ibid*

¹²⁶³ *In re Kemmler* [1890] 136 U.S. 436

¹²⁶⁴ *Hamilton v. Kentucky Distilleries & Warehouse Co.* [1919] 251 U.S. 146

¹²⁶⁵ *Kemmler*, 136 U.S. at 448-449

¹²⁶⁶ U.S. CONST., amend. 10 (“The powers not delegated to the United States by the Constitution, nor prohibited by it to the States, are reserved to the States respectively, or to the people.”)

¹²⁶⁷ *Hamilton*, 251 U.S. at 155-157

¹²⁶⁸ U.S. CONST. amend. 1

¹²⁶⁹ *Trinity*, 62 F.2d at 852

¹²⁷⁰ See *Near*, 283 U.S. at 713 (“ . . . the operation and effect of the statute in substance is that public authorities may bring the owner or publisher of a newspaper or periodical before a judge upon a charge of

ought to have been had from California's libel laws,¹²⁷¹ and indeed Shuler previously had been sued for defamation by the former mayor of Los Angeles (and was acquitted)¹²⁷² and had been found in contempt of court for broadcasting his criticism of a judge regarding a pending case.¹²⁷³

Likewise, *Near* counters the assertion in *Trinity* that progress and technology had advanced beyond the contemplation of the framers of the Constitution: Chief Justice Hughes wrote in *Near* that the importance of the immunity from previous restraint indeed had not lessened since 1776, that charges (and abuse) against public officials could not be said to be greater than at the time of the Revolution, and further, that the increased complexity of government administration in the twentieth century hazarded great risk of public malfeasance that could best be countered by a "vigilant and courageous press, especially in great cities."¹²⁷⁴ Hughes continued, "The fact that the liberty of the press may be abused by miscreant purveyors of scandal does not make any the less necessary the immunity of the press from previous restraint in dealing with official misconduct. Subsequent punishment for such abuses as may exist is the appropriate remedy, consistent with constitutional privilege."¹²⁷⁵

What the *Trinity* decision represented, rather, was the on-going realignment of the American judiciary. Decisions rejecting government regulations that the federal courts reached in earlier decades rested on legal principles like property and contract rights¹²⁷⁶ and circumspect readings of the government's power over commerce.¹²⁷⁷ These decisions reflected the strength of these social forces and the influence over them held by the judiciary itself as their arbiters, and by centres of power outside the government such as the industrial capitalists of that time.¹²⁷⁸ But since that time, the federal government's influence over commerce had grown, supported by its own burgeoning organizational strength as an

. . . publishing scandalous and defamatory matter . . . and further publication is made punishable as a contempt. This is the essence of censorship.")

¹²⁷¹ *Near*, 283 U.S. at 718-19 ("Public officers, whose character and conduct remain open to debate and free discussion in the press, find their remedies for false accusations in actions under libel laws providing for redress and punishment, and not in proceedings to restrain the publication of newspapers and periodicals.")

¹²⁷² John Samples and Paul Matzko, 'Social Media Regulation in the Public Interest: Some Lessons from History' (4 May 2020) Knight First Amendment Institute, Columbia University <<https://knightcolumbia.org/content/social-media-regulation-in-the-public-interest-some-lessons-from-history>> accessed 6 March 2022

¹²⁷³ *Trinity*, 62 F.2d at 852

¹²⁷⁴ *Near*, 283 U.S. at 719-20

¹²⁷⁵ *ibid* 720

¹²⁷⁶ See *Lochner v. New York* [1905] 198 U.S. 45 (striking down a state statute limiting bakery employee working hours)

¹²⁷⁷ See *Hammer v. Dagenhart* [1918] 247 U.S. 251 (striking down a federal law prohibiting interstate shipment of goods produced by child labour)

¹²⁷⁸ See Gaetano Mosca, *The Ruling Class* (McGraw-Hill 1939) ch 5, §§ 6 & 7

international military power, and by popular sentiment against the excesses of liberal capitalism and eventually the privations of the Great Depression. The Congress proved unable or unwilling¹²⁷⁹ to undertake the task of legislating rules for commercial industries of the scale and complexity of mass media, and so vested that power in the federal executive. The Supreme Court was no better positioned or more competent than the Congress to make the appropriate rules, and so endorsed the assignment of an increasing amount of regulatory powers to the executive branch over the 1930s and 1940s.¹²⁸⁰

The marriage between politically determined spectrum allocation and the notion of public interest gave federal regulators even greater supervisory authority as the new televisual broadcast medium arrived.¹²⁸¹ With television's tighter limit on the number of channels would come a greater risk of monopolization by less "socially desirable uses" like commercials and light entertainment.¹²⁸² Greater risk necessitated greater management, which in turn meant interfacing with "an exceptionally complex structure of institutional mediation."¹²⁸³ Accordingly, the competencies and relationships essential to that interface became increasingly vested in government administrators and corporate managers with the specific expertise for its accomplishment. The judiciary and the legislature (to say nothing of the public) would become increasingly distanced from the regulation of mass media. The Radio Commission would be replaced by the Federal Communications Commission with the enactment of the Communications Act of 1934,¹²⁸⁴ and the American approach to mass media content regulation would continue to develop through the Commission's administrative decisions.¹²⁸⁵ The Supreme Court would continue to weigh in on less technologically

¹²⁷⁹ The Congressional Record concerning the Radio Act of 1927 includes the following from a sponsor of the bill, Rep. White: "We have reached the definite conclusion that the right of all our people to enjoy this means of communication can be preserved only by the repudiation of the idea underlying the 1912 law that anyone who will may transmit and by the assertion in its stead of the doctrine that the right of the public to service is superior to the right of any individual. . . . The recent radio conference met this issue squarely. It recognized that, in the present state of scientific development, there must be a limitation upon the number of broadcasting stations, and it recommended that licenses should be issued only to those stations whose operation would render a benefit to the public." 67 CONG. REC. 5479

¹²⁸⁰ See, e.g., *Wickard v. Filburn* [1942] 317 U.S. 111 (approving the reach of federal regulation even to wholly local activity due to aggregate effect on interstate commerce)

¹²⁸¹ See Lee M Mitchell, 'Who Watches the Watchdog?: The Crisis of Public Confidence' in William L Rivers & Michael J Nyhan (eds), *Aspen Notebook on Government and the Media* (Praeger London 1973) 144, 146; these purposive justifications for the regulator would be adopted in Europe as well, as in Article 10's licensing provisos and in the right to receive information: see also A Nichol, G Millar, and A Sharland, *Media Law and Human Rights* (2nd edn, OUP 2009) 195-96, 16, 23

¹²⁸² Dan Lacy, 'Rules for the Regulators: A Rationale for Government's Role' in William L Rivers & Michael J Nyhan (eds), *Aspen Notebook on Government and the Media* (Praeger London 1973) 93, 97, 114

¹²⁸³ Dan Lacy, quoted in Roger Fisher, 'Press Rights vs. Press Responsibilities: The Issues of the First Amendment' in William L. Rivers & Michael J. Nyhan (eds), *Aspen Notebook on Government and the Media* (Praeger London 1973) 51, 67; see also Lacy (n 1282) 93

¹²⁸⁴ Communications Act of 1934, 73 PUB.L. 416

¹²⁸⁵ See, e.g., *Young People's Association for the Propagation of the Gospel*, [1938] 6 F.C.C. 178, 180 (denying as not "serving the public interest" a broadcasting license application on the grounds the

mediated free speech cases over which its judgements still held substantial sway, involving state laws on pamphlet distribution and picketing, for example.¹²⁸⁶

Moves Toward Markets as Arbiters

In this section, I discuss how changes in the structure of American television—particularly its progression from vertically integrated over-the-air broadcasters to more horizontal cable platforms—interacted with existing regulatory structures and legal rationales. During this phase, the FCC and the Supreme Court had to reckon with these changes and decide on alternative ways (like relying on markets) to achieving public aims.

Content regulation for American broadcast television and radio bearing on political news chiefly developed through the Commission’s fairness doctrine—a set of broadcasting policies which, as discussed in Part Five, have received renewed attention as a potential wellspring for social media content regulation. The Supreme Court’s 1969 validation of the fairness doctrine marked its continued endorsement the prevailing picture of radio and television mass media as “a prime source of national cohesion and news.”¹²⁸⁷ Furthermore, the *Red Lion* Court articulated an expansive vision of the First Amendment—one with implications for subsequent decisions on content regulation, as former FCC Commissioner Cox explains:

“[W]hen the Court confirmed constitutionality of the Fairness Doctrine and used, as it often does, the broad language that has implications elsewhere that at least that group of Justices felt that the First Amendment did mean something more than just preservation against government or state action of the right of the individual to speak. Now of course the decision was in a context of licensed media—media in which the licensees enjoy the benefit of a government policy Maybe that’s all the court was talking about, but it does lend some credence to the view that if Congress were to take affirmative action, effective or not, to try to promote greater freedom in the print media, the Court might not regard that as violating the First Amendment. It might say that it is really promoting it. The issue is not so much the right of the

applicant’s proposed religious programming would only include programs that coincided with the applicant’s religious beliefs)

¹²⁸⁶ See, e.g., *Schneider v. State* [1939] 308 U.S. 147 and *Thornhill v. Alabama* [1940] 310 U.S. 88

¹²⁸⁷ *Red Lion*, 395 U.S. 386; Denis McQuail, *Media Accountability and Freedom of Publication* (OUP 2007) 6 (describing the public’s expectations at the time of the mass media in this role: that the citizens can be informed and active participants in elections and be able to hold governments “continuously and publicly accountable for their actions and omissions” with the help of the media.)

broadcaster to speak as it is the right of the public to hear and to be informed."¹²⁸⁸

The Supreme Court's rationale in *Red Lion* indeed took place in the context of licensed media, but nevertheless was recognized by observers for its potential application to other media where the right of the public to be informed could be implicated.¹²⁸⁹ Similar concerns expressed over the reach, influence, and power of social media platforms acting in the public sphere have reinvigorated American debates about the "legitimacy of government's affirmative duty to protect positive freedoms."¹²⁹⁰

Even by the time of *Red Lion*, however, the fairness doctrine had become somewhat of a morass, from the perspectives of the public, the broadcast media, and the regulators alike.¹²⁹¹ Citizens' groups sought more robust formats to address controversial advertising like that of cigarette makers and car manufacturers, the media bemoaned advertisers being run off (keeping in mind that the U.S. broadcast model is chiefly commercial, i.e., paid for by advertising), and the regulators were loath to become too involved in second-guessing licensees' business decisions.¹²⁹² Technological advancements, simply by adding more channels, continued to complicate the role of the FCC by lessening the implication that any particular licensing decision would have an outsized effect on the public interest. While the drafters of the 1927 Radio Act agreed to the radio industry's suggested 89 different practicable wavelengths, they were not completely naïve to the possibility of increasing the size of the broadcasting band, even with a limited understanding of the then-new technology.¹²⁹³ The number of workable stations did increase dramatically with advances like FM radio in the 1940s¹²⁹⁴ and the development of UHF television in the 1960s.¹²⁹⁵ The

¹²⁸⁸ Quoted in Fisher (n 1283) 73

¹²⁸⁹ *ibid*

¹²⁹⁰ Victor Pickard, 'The Fairness Doctrine won't solve our problems – but it can foster needed debate' (04 February, 2021) *The Washington Post* <<https://www.washingtonpost.com/outlook/2021/02/04/fairness-doctrine-wont-solve-our-problems-it-can-foster-needed-debate/>> accessed 05 July 2022; see also Philip M. Napoli, 'Lessons for social media from the Fairness Doctrine,' (13 August 2020) *Columbia Journalism Review* <<https://www.cjr.org/opinion/social-media-fairness-doctrine.php>> accessed 05 July 2022

¹²⁹¹ Branscomb (n 1244) 122-23

¹²⁹² *ibid*

¹²⁹³ Charlotte Twight, 'What Congressmen Knew and When They Knew It: Further Evidence on the Origins of U.S. Broadcasting Regulation' (June 1998) 95(3) *Public Choice* 247, 258-59

¹²⁹⁴ See Federal Communications Commission, 'History of Commercial Radio' (April 12, 2021) <<https://www.fcc.gov/media/radio/history-of-commercial-radio>> accessed 05 July 2022

¹²⁹⁵ See Ken Gilmore, 'Must We Have UHF-TV?' (May 1962) *Popular Electronics*, available at <https://www.rfcafe.com/references/popular-electronics/must-we-have-uhf-tv-may-1962-popular-electronics.htm>

growth of cable television in the 1970s and 1980s further eroded the scarcity basis which had been fundamental to the regulatory approach for broadcasting in the United States.¹²⁹⁶

By 1940, the U.S. Federal Communications Commission (“FCC”) had laid out the full statement of its expectations for responsible broadcasters, which came to be known as the “fairness doctrine” over the following decades of its application¹²⁹⁷ to radio and later television:

“[With] the limitations in frequencies inherent in the nature of radio, the public interest can never be served by a dedication of any broadcast facility to the support of [the broadcaster’s] own partisan ends. Radio can serve as an instrument of democracy only when devoted to the communication of information and the exchange of ideas fairly and objectively presented. A truly free radio cannot be used to advocate the causes of the licensee. . . . Freedom of speech on the radio must be broad enough to provide full and equal opportunity for the presentation to the public of all sides of public issues. Indeed, as one licensed to operate in a public domain the licensee has assumed the obligation of presenting all sides of important public questions, fairly, objectively and without bias. The public interest—not the private—is paramount. These requirements are inherent in the conception of public interest set up by the Communications Act as the criterion of regulation.”¹²⁹⁸

The public interest standard continued forward from the radio era into the era of broadcast television as the basis for requirements upon licensees to provide opportunities for the discussion of both sides of controversial public issues.¹²⁹⁹ The Congress and FCC further developed the fairness doctrine into the 1960s, particularly with regard to licensees affording equal time to political candidates making campaign appearances.¹³⁰⁰ In 1969, the Supreme Court upheld the FCC’s promulgation of additional rules complementary to its express ambit, citing the Commission’s pursuit of the public interest.¹³⁰¹ The *Red Lion* decision cited the “differences in the characteristics of new media” to justify differences in First Amendment standards applied to broadcasters.¹³⁰² Those characteristics included the

¹²⁹⁶ McQuail (n 1287) 260; Lacy (n 1282) 97 & 114

¹²⁹⁷ See generally *Red Lion Broadcasting Co., Inc. v. F.C.C.*, 395 U.S. 367 (1969)

¹²⁹⁸ *Mayflower Broadcasting Corp. and The Yankee Network, Inc.* [1940] 8 F.C.C. 333, 339-40

¹²⁹⁹ *Red Lion*, 395 U.S. at 382

¹³⁰⁰ *ibid* 380-86

¹³⁰¹ *ibid*

¹³⁰² *ibid* 386

reach of radio signals,¹³⁰³ the apportionment of the spectrum,¹³⁰⁴ and technological advancements permitting new uses and greater efficiency.¹³⁰⁵

The Court in *Red Lion* also extended the government's active role in the broadcast media's meeting public expectations—as guarantor of “the right of the public to receive suitable access to social, political, esthetic [sic], moral, and other ideas and experiences.”¹³⁰⁶ This formulation is much closer in letter and spirit to Article 10 ECHR than to the First Amendment of the U.S. Constitution, and reflects an on-going debate in the U.S. over the notion of positive obligations on the government to secure free expression and a free press. Article 10(1) provides for the right to impart and receive information and explicitly allows the imposition of licensing schemes for broadcasting and television.¹³⁰⁷ The European Court has made clear that such licensure may be conditioned on public interest considerations above and beyond ‘technical aspects’¹³⁰⁸ albeit within the requirements of Article 10(2).¹³⁰⁹ Ofcom's Broadcasting Code sets expectations of “due impartiality” and “due accuracy” for its licensed news broadcasters¹³¹⁰ pursuant to these principles, for example. Distinguishing between the American First Amendment and European Article 10 approaches, Murray suggests certain legal traditions and historical experiences in Europe came to be reflected in national differences as the reach of the Internet grew to address international audiences.¹³¹¹ The exceptions in Article 10(2)—while prescriptive of some of the same constraints espoused by American constitutional law with regard to not being vague (having to be provided by law) and being subject to a level of judicial scrutiny (being necessary and proportionate)—permit a more “extensive” array of national interests to be served by restrictions on the freedom of expression in comparison with U.S. law.¹³¹²

Whether such policies, as applied to the new media of the Internet era, would more or less closely resemble ‘public interest’-driven regulations like the Fairness Doctrine for broadcasters of previous eras have been the subject of debate and criticism by academics and politicians alike. Napoli has called for a public interest orientation in social media

¹³⁰³ *Red Lion*, 395 U.S. at 387-88

¹³⁰⁴ *ibid* 388-391

¹³⁰⁵ *ibid* 396-400

¹³⁰⁶ *ibid* 390

¹³⁰⁷ European Convention on Human Rights, Article 10

¹³⁰⁸ Nichol, Millar, and Sharland (n 1281) 24 (quoting *Informationsverein Lentia and others v Austria* (1993) 17 EHRR 93, at para 32)

¹³⁰⁹ *ibid* 195-196

¹³¹⁰ Ofcom Broadcasting Code, ‘Section five: Due impartiality and due accuracy’ (05 January 2021)

<<https://www.ofcom.org.uk/tv-radio-and-on-demand/broadcast-codes/broadcast-code/section-five-due-impartiality-accuracy>> accessed 05 July 2022

¹³¹¹ Andrew Murray, *Information Technology Law: The Law & Society* (2nd edn, OUP 2013) 114-17 (in particular, the experience of World War II)

¹³¹² *ibid*

platform governance, arguing that the key difference between traditional media and social media is the lack of those public interest frameworks that characterized the professional norms and regulatory context of the former.¹³¹³ Napoli's reasoning focuses on the displacement of 'old' media gatekeepers by irresponsible or unaccountable institutions, particularly where the gatekeeping function is accomplished by algorithms.¹³¹⁴ According to this perspective, the notion that social media platforms allow users to access, create, and interact with more news and information than ever before is actually *disempowering*, because the platforms are insulated from the stakeholder pressures that upheld public service and social responsibility functions in traditional media.¹³¹⁵

Arguably, social media presents similar characteristics to broadcasting in terms of reach, new uses, and greater efficiency, but lacks the essential characteristics of licensed media (like spectrum scarcity) that justified government intrusion in *Red Lion*. Although Napoli acknowledges the probable inapplicability of Fairness Doctrine-style regulation to social media,¹³¹⁶ he suggests envisioning the user data aggregated by the platforms as a sort of public resource akin to the broadcast spectrum.¹³¹⁷ Pickard offers a similar account of new media "run-amok,"¹³¹⁸ particularly with regard to the purported monopoly power of companies like Facebook.¹³¹⁹ This perspective likewise recognizes the historical role of professionalization in balancing public service and profit motives in traditional media as compared to social media, but also suggests that previously existing structural flaws in the 'market' for journalism are reproduced in the Internet age.¹³²⁰ Pickard contends that the public interest would be best served by public ownership and democratically governed media system alternatives.¹³²¹

Cable television carried with it the additional complication of delivering both broadcast programming, which was subject to broadcast regulation, and also paid subscriber content, which arguably cannot be so casually encountered as broadcast media and

¹³¹³ Philip M Napoli, *Social Media and the Public Interest: Media Regulation in the Disinformation Age* (2019, Columbia UP) 15-16

¹³¹⁴ *ibid* 17-18

¹³¹⁵ *ibid* 17

¹³¹⁶ Philip M Napoli, 'Lessons for social media from the Fairness Doctrine,' (13 August 2020) *Columbia Journalism Review* <<https://www.cjr.org/opinion/social-media-fairness-doctrine.php>> accessed 05 July 2022

¹³¹⁷ Philip M Napoli, 'What Would Facebook Regulation Look Like? Start With the FCC' (*Wired*, 04 October 2019) <<https://www.wired.com/story/what-would-facebook-regulation-look-like-start-with-the-fcc/>> accessed 26 September 2022

¹³¹⁸ Victor Pickard, 'The Fairness Doctrine won't solve our problems – but it can foster needed debate' (04 February, 2021) *The Washington Post* <<https://www.washingtonpost.com/outlook/2021/02/04/fairness-doctrine-wont-solve-our-problems-it-can-foster-needed-debate/>> accessed 05 July 2022

¹³¹⁹ Victor Pickard, *Democracy without Journalism? Confronting the Misinformation Society* (2019 Oxford UP)

¹³²⁰ *ibid*

¹³²¹ *ibid*

accordingly ought not be so readily regulated.¹³²² As the market for cable programming developed, the FCC established various rules for cable television over the 1970s across a range of content issues, applying broadcast regulation concepts like the fairness doctrine through the Commission's general authority over cable.¹³²³ The dual classification of the cable operators as acting similarly to "broadcasters on their program-origination channel" and like "a common carrier on the public access and leased channels" necessitated different regulatory approaches to each role, generating "confusion over the responsibility for content."¹³²⁴ By the late 1980s, the FCC abandoned the fairness doctrine,¹³²⁵ but other broadcasting rules (such as the requirement that qualified candidates for public office be given equal opportunities to make television appearances) have continued to be applied to cable television in circumstances where system operators exercise exclusive control over programming.¹³²⁶

A pair of Supreme Court cases¹³²⁷ in the mid-1990s highlighted the ways in which the widespread adoption of cable systems had shifted the social interests at stake in American telecommunications. Those particular interests would take on additional significance in the same decade as more people began using the Internet. The *Turner* cases arose after the Congress enacted the 1992 Cable Act,¹³²⁸ which *inter alia* required cable television operators to provide carriage to commercial broadcast stations and non-commercial, educational broadcasters operating in the same local television market. Several cable programmers and operators challenged the constitutionality of these 'must-carry' provisions under the First Amendment.¹³²⁹ The Supreme Court recognized the technological distinctions between cable and over-the-air broadcasting as decidedly favouring cable, as cable "eliminates the signal interference sometimes encountered . . . [and] is capable of transmitting many more channels."¹³³⁰ The Court thereby reasoned that the regulatory rationale in broadcasting, justified by scarcity—the physical limitation of the medium—would be inapposite to cable

¹³²² Lacy (n 1282) 112 (as discussed by Commissioner Cox in their conversation)

¹³²³ Henry Geller and Donna Lampert, 'Cable, Content Regulation and the First Amendment' (1983) 32(3) Catholic Univ L Rev 603, 605-07

¹³²⁴ Branscomb (n 1244) 124; see also Lacy (n 1282) 113-15 (discussing what non-regulatory relationship government might have to paid television in order to promote cultural and educational programming, quality)

¹³²⁵ Federal Communications Commission, 'FCC Chairman Genachowski continues regulatory reform to ease burden on businesses; announces elimination of 83 outdated rules' (22 August 2011) DOC 309224A1 <<https://docs.fcc.gov/public/attachments/DOC-309224A1.pdf>> accessed 05 July 2022 (formally striking the dead letter regulation)

¹³²⁶ Federal Communications Commission, 'Program Content Regulations' (29 September 2017) <<https://www.fcc.gov/media/program-content-regulations>> accessed 05 July 2022; also 47 U.S. CODE § 315 of the Communications Act of 1934

¹³²⁷ *Turner Broadcasting System, Inc. v. FCC* [1994] 512 U.S. 622 ("Turner I"); *Turner Broadcasting System, Inc. v. FCC* [1997] 520 U.S. 180 ("Turner II")

¹³²⁸ Cable Television Consumer Protection and Competition Act of 1992, 102 PUB. L. 385

¹³²⁹ See *Turner I*, 512 U.S. at 634-36

¹³³⁰ *ibid* 628

transmission, which even in the early 1990s heralded a future with “no practical limitation on the number of speakers who may use [it].”¹³³¹

While describing the function of the cable system “in essence, as a conduit for the speech of others,” the Court unambiguously stated that cable operations are First Amendment activities, and that the must-carry provisions indeed regulated speech and should therefore satisfy more than mere ‘rational basis’ constitutional scrutiny.¹³³² The *Turner* Court thus evaluated whether the regulations advanced an “important” government interest and did not burden “substantially more speech than necessary” to further that interest—actually three interests in this case: “(1) preserving the benefits of free, over-the-air local broadcast television, (2) promoting the widespread dissemination of information from a multiplicity of sources, and (3) promoting fair competition in the market for television programming.”¹³³³

The *Turner* Court also compared cable television to newspapers, both in terms of market characteristics and technological differences.¹³³⁴ While either medium might tend towards a local monopoly in a given area, the newspaper lacks the direct physical connection to the subscriber that allows the cable operator to exclude access to competing speakers “simply by virtue of its ownership of the essential pathway for cable speech.”¹³³⁵ The Court therefore determined that unlike an unconstitutional ‘right-of-reply’ provision affecting newspapers, the must-carry provisions for cable were not an unjustifiable intrusion on operators’ editorial control, but instead aimed to “ensure that private interests not restrict, through physical control of a critical pathway of communication, the free flow of information and ideas.”¹³³⁶ A plurality of the Justices decided that objective was sufficiently important and that the must-carry provision operated without regard to program content, so the first *Turner* case was remanded for additional fact-finding to determine if the provisions fit the objective.¹³³⁷

The Court continued to emphasize the social interests at stake in the structure of the television market in the second *Turner* case: ensuring public access to “a multiplicity of

¹³³¹ *Turner I*, 512 U.S. at 637-39

¹³³² *ibid* 636-37 (“By requiring cable systems to set aside a portion of their channels for local broadcasters, the must-carry rules regulate cable speech in two respects: The rules reduce the number of channels over which cable operators exercise unfettered control, and they render it more difficult for cable programmers to compete for carriage on the limited channels remaining.”)

¹³³³ *ibid* 662 (applying the intermediate level of scrutiny applicable to “content-neutral” restrictions)

¹³³⁴ *ibid* 656-657

¹³³⁵ *ibid* 656 (describing this as “gatekeeper” control, done at the “mere flick of the switch”)

¹³³⁶ *ibid* 653-57 (discussing *Miami Herald Publishing Co. v. Tornillo*, 418 U.S. 241 (1974))

¹³³⁷ *ibid* 661-74 (J. Blackmun concurring, J. Stevens concurring in part and concurring in the judgment)

information sources,”¹³³⁸ addressing the shift in market share from broadcast to cable and concomitant concentration and integration,¹³³⁹ promoting “the widest possible dissemination of information from diverse and antagonistic sources” as “essential to the welfare of the public”¹³⁴⁰, and preventing “a reduction in the number of media voices available to consumers.”¹³⁴¹ Based on the expanded record from the trial on remand, the Court deferred to the determination by the Congress that broadcast stations were in jeopardy¹³⁴² and that Congress reached a reasonable legislative conclusion in the form of the must-carry provisions.¹³⁴³ Since the number of broadcast stations added to cable systems after the must-carry rules became effective presumably equalled the number of stations that would have been denied carriage in the absence of the must-carry requirements, the Court concluded the provisions were “narrowly tailored to preserve a multiplicity of broadcast stations for the 40 percent of American households without cable.”¹³⁴⁴

However, the dissenting Justices across both *Turner* cases remained unconvinced that the must-carry provisions were content-neutral (rather, that they reflected a preference for “local, educational, or informational” programming, however benign that might be)¹³⁴⁵ or that the assumption “that most adverse carriage decisions are anticompetitively motivated, and that must-carry is therefore a measured response to a problem of anticompetitive behaviour,”¹³⁴⁶ was a fair one. In discussing less-restrictive alternatives (like common carrier rules, subsidies, and vertical integration controls)¹³⁴⁷ and technical advances (new transmission media like satellite, expanded capacity through fiber-optic networks),¹³⁴⁸ the *Turner* dissenters appear to have done a better job tracking trends in communications than the 1992 Cable Act Congress had, despite the deference shown that body in “judgments about the ‘interaction of industries undergoing rapid economic and technological change.’”¹³⁴⁹

The Court in *Turner I* makes clear at least one line of reasoning that bears on whether social media platforms engage in speech (notwithstanding the legal distinction provided by Section 230 that says they are not publishers or speakers of user-generated content):

¹³³⁸ *Turner II*, 520 U.S. at 190

¹³³⁹ *ibid* 191

¹³⁴⁰ *ibid* 192

¹³⁴¹ *ibid* 193

¹³⁴² *ibid* 195-96

¹³⁴³ *ibid* 209-11

¹³⁴⁴ *ibid* 215-16

¹³⁴⁵ *Turner I*, 512 U.S. at 675-82; *Turner II*, 520 U.S. at 230 & 233-34

¹³⁴⁶ *Turner I*, 512 U.S. at 683-84; *Turner II*, 520 U.S. at 232-33

¹³⁴⁷ *Turner I*, 512 U.S. at 684; *Turner II*, 520 U.S. at 238

¹³⁴⁸ *Turner I*, 512 U.S. at 684; *Turner II*, 520 U.S. at 245

¹³⁴⁹ *Turner II*, 520 U.S. at 246

“There can be no disagreement on an initial premise: Cable programmers and cable operators engage in and transmit speech, and they are entitled to the protection of the speech and press provisions of the First Amendment. Through ‘original programming or by exercising editorial discretion over which stations or programs to include in its repertoire,’ cable programmers and operators ‘see[k] to communicate messages on a wide variety of topics and in a wide variety of formats.’”¹³⁵⁰

Justice O’Connor agreed:

“Though cable operators do not actually originate most of the programming they show, the Court correctly holds that they are, for First Amendment purposes, speakers. Selecting which speech to retransmit is, as we know from the example of publishing houses, movie theaters, bookstores, and Reader’s Digest, no less communication than is creating the speech in the first place.”¹³⁵¹

In the absence of Section 230, the *Turner* Court might have reached similar conclusions about social media platforms, viewing their business as fundamentally similar to that of cable operators, and accordingly contemplate similar treatment for the platforms. That treatment might look something like must-carry provisions for user accounts representing regulated broadcast journalism outlets and public officials (essentially the requirement of SB 7072’s §§ 501.2041 (2)(h),(j)¹³⁵²), or even some of the less-restrictive alternatives contemplated by the *Turner* dissenters like common carriage or vertical integration controls—potentially in the form of platform interoperability requirements as discussed in Part Five.¹³⁵³

¹³⁵⁰ *Turner I*, 512 U.S. at 636-37 (internal citations omitted)

¹³⁵¹ *ibid* 675 (J. O’Connor, concurring in part and dissenting in part)

¹³⁵² Part V, ‘Public Policy and the Public Sphere’ – American Perspectives

¹³⁵³ Part V, ‘Analysing Allocation Decisions’ – Legislative Control

Bibliography

Primary Sources

Abrams v. U.S. [1919] 250 U.S. 616

Ashcroft v. American Civil Liberties Union [2004] 542 U.S. 656

Brandenburg v. Ohio [1969] 395 U.S. 444

Buckley v. Valeo [1976] 424 U.S. 1

Gitlow v. New York [1925] 268 U.S. 652

Gonzalez v. Google, LLC, No. 21-1333, *cert. granted* (03 October 2022)

Hamilton v. Kentucky Distilleries & Warehouse Co. [1919] 251 U.S. 146

Hammer v. Dagenhart [1918] 247 U.S. 251

Hoover, Sec'y Comm. v. Intercity Radio Co., Inc. [1923] 52 App.D.C. 339

Hurley v. Irish-American Gay, Lesbian and Bisexual Group of Boston [1995] 515 U.S. 557

In re Kemmler [1890] 136 U.S. 436

Lochner v. New York [1905] 198 U.S. 45

Lovell v. Griffin [1938] 303 U.S. 444

Mayflower Broadcasting Corp. and The Yankee Network, Inc. [1940] 8 F.C.C. 333

Miami Herald Publishing Co. v. Tornillo [1974] 418 U.S. 241

Near v. Minnesota [1931] 283 U.S. 697

NetChoice, LLC v. Paxton, No. 21-51178 (5th Cir. 2022)

NetChoice, LLC v. Att'y Gen. of Fla., No. 21-12355 (11th Cir. 2022)

Att'y Gen. of Fla. v. NetChoice, LLC, No. ___, *Petition for a Writ of Certiorari* (21 September 2022)

NetChoice, LLC v. Moody, No. 21-12355 (11th Cir., 23 May 2022)

Pierce v. United States [1920] 252 U.S. 239

Red Lion Broadcasting Co., Inc. v. F.C.C., 395 U.S. 367 (1969)

Reno v. American Civil Liberties Union [1997] 521 U.S. 844

Schaefer v. United States [1920] 251 U.S. 466

Schenck v. U.S. [1919] 249 U.S. 47

Schneider v. State [1939] 308 U.S. 147

Thornhill v. Alabama [1940] 310 U.S. 88

Trinity Methodist Church, South v. Federal Radio Comm'n [1932] 62 F.2d 850

Turner Broadcasting System, Inc. v. Federal communications Commission [1997] 520 U.S. 180

Turner Broadcasting System, Inc. v. Federal Communications Commission [1994] 512 U.S. 622

Whitney v. California [1927] 274 U.S. 357

Wickard v. Filburn [1942] 317 U.S. 111

Young People's Association for the Propagation of the Gospel, [1938] 6 F.C.C. 178

U.S. v. Zenith Radio Corporation [1926] 12 F.2d 614

U.S. CONST. amend. 1

U.S. CONST., art. 1, s. 8

ACCESS Act of 2021, H.R. 3849 (117th Congress, 1st Session, 11 June 2021)

Cable Television Consumer Protection and Competition Act of 1992, 102 PUB. L. 385

Communications Act of 1934, 73 PUB.L. 416

Communications Decency Act of 1996, 47 U.S. CODE § 230

Digital Millennium Copyright Act of 1998, 105 PUB. L. 304

Telecommunications Act of 1996, 104 PUB. L. 104

Ending Support for Internet Censorship Act, S. 1914 (116th Congress, 1st session, 19 June 2019)

Espionage Act of 1917, 65 PUB. L. 24

Amendment of 1918 to the Espionage Act of 1917, 65 PUB. L. 150

Radio Act of 1912, 37 STAT. 302

Radio Act of 1927, 44 STAT. 1162

FED. R. EVID. 403, 404(a)

Congressional Record (4 August 1995) 141 CONG. REC. H8468-70

Congressional Record (16 March 1926) 67 CONG. REC. 5479

Senate Bill No. 7072, Laws of Florida, Chapter 2021-32 (effective 01 July 2021)

House Bill 20, TEXAS BUSINESS & COMMERCE CODE, Subtitle C, Title 5, Chapter 120, and TEXAS CIVIL PRACTICE AND REMEDIES CODE, Title 6, Chapter 143A (effective 09 September 2021)

Electronic Commerce (EC Directive) Regulations 2002

Ofcom Broadcasting Code, 'Section five: Due impartiality and due accuracy' (05 January 2021) <<https://www.ofcom.org.uk/tv-radio-and-on-demand/broadcast-codes/broadcast-code/section-five-due-impartiality-accuracy>> accessed 05 July 2022

Online Safety Bill, Bill 285 2021-22 (as introduced, 17 March 2022), Bill 121 2022-23 (as amended in Public Bill Committee, 28 June 2022)

<<https://publications.parliament.uk/pa/bills/cbill/58-03/0121/220121.pdf>>

Online Safety Bill, Bill 220 2022-23 (as amended on re-committal, 19 December 2022)

§ 86 StGB [Germany, 1998] (translation available at <https://www.gesetze-im-internet.de/englisch_stgb/englisch_stgb.html>) accessed 25 February 2020

§ 130 StGB [Germany, 1998] (translation available at <https://www.gesetze-im-internet.de/englisch_stgb/englisch_stgb.html>) accessed 25 February 2020

European Convention on Human Rights, Article 10

Commission, 'Tackling online disinformation: a European Approach' COM (2018) 236 final, 26 April 2018)

Commission, 'European Commission Guidance on Strengthening the Code of Practice on Disinformation' COM (2021) 262 final, 26 May 2021

Commission, 'Online Platforms and the Digital Single Market: Opportunities and Challenges for Europe' COM (2016) 288 final, 25 May 2016

Commission, 'EU Code of Practice on Disinformation' (2018)
<<https://ec.europa.eu/newsroom/dae/redirection/document/87534>>

Commission, 'The Strengthened Code of Practice on Disinformation 2022'
<<https://ec.europa.eu/newsroom/dae/redirection/document/87585>>

Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market

Regulation 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act)

Secondary Sources

12 Angry Men, dir. by Sidney Lumet (United Artists, 1957)

Acemoglu D and Restrepo P, 'Tasks, Automation, and the Rise in US Wage Inequality' (*National Bureau of Economic Research*, June 2021) NBER Working Paper 28920
<<https://www.nber.org/papers/w28920>> accessed 01 December 2022

Ajzen I and Fishbein M, 'Attitudes and the Attitude-Behavior Relation: Reasoned and Automatic Processes' (2000) 11 *European Review of Social Psychology* 1

Alabama Historical Radio Society, 'WSY – First Radio Station in Alabama' (2020)
<<https://alhrs.org/wsy/>> accessed 6 March 2022

Al Jazeera Investigations, 'The Labour Files' (Al Jazeera Investigations, 23 September 2022)
<<https://www.ajunit.com/investigation/the-labour-files/>> accessed 01 November 2022

Allcott H and Gentzkow M, 'Social Media and Fake News in the 2016 Election' (2017) 31(2) *J Econ Perspectives* 211

Allcott H and others, 'The Welfare Effects of Social Media' (2019) National Bureau of Economic Research Working Paper 25514 <<http://www.nber.org/papers/w25514>> accessed 28 July 2022

Amazon Mechanical Turk (2018) <<https://www.mturk.com>> last accessed 4 May 2021

Anderson C A, Lepper M R, and Ross L, 'Perseverance of Social Theories: The Role of Explanation in the Persistence of Discredited Information' (1980) 39(6) *J Personality and Social Psychology* 1037

Anspach N M, Jennings J T, and Arceneaux K, 'A little bit of knowledge: Facebook's News Feed and self-perceptions of knowledge' (January-March 2019) *Research and Politics* 1

Article One Advisors, 'Assessing the Human Rights Impact of the Facebook Platform in Sri Lanka' (2018) <https://www.articleoneadvisors.com/s/Sri-Lanka-HRIA_-Executive-Summary_FINAL.pdf> accessed 28 July 2022

Asch S E, 'Studies of Independence and Conformity: I. A Minority of One Against a Unanimous Majority' (1956) 70(9, Whole No. 416) *Psychological Monographs* 1

Atlantic Council, 'Atlantic Council's Digital Forensic Research Lab Partners with Facebook to Combat Disinformation in Democratic Elections' (17 May 2018)

<<https://www.atlanticcouncil.org/news/press-releases/atlantic-councils-digital-forensic-research-lab-partners-with-facebook/>> accessed 28 July 2022

Balkin J M, 'Digital speech and democratic culture: A theory of freedom of expression for the information society' (2004) 79 NYU L Rev 1

Barlow J P, 'A Declaration of the Independence of Cyberspace' (Electronic Frontier Foundation, 08 February 1996) <<https://www.eff.org/cyberspace-independence>> accessed 28 July 2022

Bastani A, "'It's going to be a long night" – How Members of Labour's Senior Management Team Campaigned to Lose' (Novara Media, 12 April 2020)

<<https://novaramedia.com/2020/04/12/its-going-to-be-a-long-night-how-members-of-labours-senior-management-campaigned-to-lose/>> accessed 01 November 2022

BBC, 'General election 2019: No apology from Jeremy Corbyn over Labour anti-Semitism claims' (bbc.co.uk, 26 November 2019) <<https://www.bbc.co.uk/news/election-2019-50564965>> accessed 29 October 2022

BBC, 'Nick Clegg gets bigger role at Facebook owner' (BBC.com, 16 Feb 2022) <<https://www.bbc.com/news/uk-60410636>> accessed 01 April 2023

BBC Panorama, 'Is Labour Anti-Semitic?' (10 July 2019)

<<https://www.bbc.co.uk/programmes/m0006p8c>> accessed 29 October 2022

Bensman M R, 'Intercity Radio Company Case' in Donald G Godfrey & Frederic A Leigh (eds.) *Historical Dictionary of American Radio* (Greenwood Press 1998)

Bergami M and Bagozzi R P, 'Self-categorization, affective commitment and group self-esteem as distinct aspects of social identity in the organization' (2000) 39 British J Social Psychology 555

Bernal P, 'Fakebook: why Facebook makes the fake news problem inevitable' (2018) 69(4) Northern Ireland Legal Quarterly 513

Bernal P, 'The push to ban Nazis on Twitter can just as easily affect the left – as the suspension of a Corbin supporter proves' (*Independent*, 05 March 2019)

<https://www.independent.co.uk/voices/twitter-ban-nazi-suspend-rachael-swindon-labour-corbyn-a8808156.html>> accessed 28 July 2022

Bernstein C, 'The CIA and the Media' (Rolling Stone, 20 October 1977)
<<https://www.carlbernstein.com/the-cia-and-the-media-rolling-stone-10-20-1977>>
accessed 28 July 2022

Bertot J C, Jaeger P T, and Grimes J M, 'Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies' (2010) 27 Government Information Quarterly 264

Bertot J C, Jaeger P T, and Hansen D, 'The impact of policies on government social media usage: Issues, challenges, and recommendations' (2012) 29 Government Information Quarterly 30

Bertrand N, 'Hunter Biden story is disinfo, dozens of former intel officials say' (Politico, 19 October 2020) <<https://www.politico.com/news/2020/10/19/hunter-biden-story-russian-disinfo-430276>> accessed 01 November 2022

Berzina K and Soula E, 'Conceptualizing Foreign Interference in Europe' (*Alliance for Securing Democracy*, 19 March 2020) <<https://securingdemocracy.gmfus.org/wp-content/uploads/2020/03/Conceptualizing-Foreign-Interference-in-Europe.pdf>>
accessed 28 July 2022

Biddle S, 'Facebook suspended a Latin American news network and gave three different reasons why' (*The Intercept*, 20 August, 2018)
<<https://theintercept.com/2018/08/20/facebook-suspended-a-latin-american-news-network-and-gave-three-different-reasons-why>> accessed 28 July 2022

Bizer G Y, Larsen J T, and Petty R E, 'Exploring the Valence-Framing Effect: Negative Framing Enhances Attitude Strength' (2011) 32 (1) Political Psychology 59

Blackburn R, 'What is forensic psychology?' (1996) 1 Legal and Criminological Psychology 3

Bó E, 'Regulatory Capture: A Review' (2006) 22(2) Oxford Review of Economic Policy 203

Boltanski L and Chiapello E, 'The New Spirit of Capitalism' (2005) 18 Int J Polit Cult Soc 161

- Boyd D, 'You Think You Want Media Literacy... Do You?' (*Data & Society: Points*, 9 March 2018) <<https://points.datasociety.net/you-think-you-want-media-literacy-do-you-7cad6af18ec2>>
- Bradshaw A S and others, 'Propagandizing anti-vaccination: Analysis of *Vaccines Revealed* documentary series' (2020) 38 Vaccine 2058
- Branscomb A, 'Citizen Access to the Media' in William L Rivers & Michael J Nyhan (eds), *Aspen Notebook on Government and the Media* (Praeger London 1973) 120
- Broderick R and Darmanin J, 'The "Yellow Vest" Riots in France Are What Happens When Facebook Gets Involved With Local News' (BuzzFeed, 06 December 2018) <<https://www.buzzfeednews.com/article/ryanhatesthis/france-paris-yellow-jackets-facebook>> accessed 28 July 2022
- Burke J, *The Day the Universe Changed* (rev edn, Little, Brown 1995)
- Cadwalladr C, 'Parliament seizes cache of Facebook internal papers' (*The Guardian*, 24 November 2018) <<https://www.theguardian.com/technology/2018/nov/24/mps-seize-cache-facebook-internal-papers>> accessed 28 July 2022
- Cadwalladr C and Graham-Harrison E, 'Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach' (*The Observer*, 17 March 2018) <<https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>> accessed 28 August 2022
- Cameron K A, Jacks J Z, and O'Brien M E, 'An Experimental Examination of Strategies for Resisting Persuasion' (2002) 7 Current Research in Social Psychology 205
- Cammaerts B, 'Radical pluralism and free speech online: the case of North Belgian extreme right discourses' (2009) 12(6) Int'l J Cultural Studies 555
- Caplan R, 'Content or Context Moderation?: Artisanal, Community-Reliant, and Industrial Approaches' (*Data & Society*, 14 November 2018) <<https://datasociety.net/library/content-or-context-moderation/>> accessed 28 July 2022
- Casey J P, Vanman E J, and Barlow F K, 'Empathetic Conservatives and Moralizing Liberals: Political Intergroup Empathy Varies by Political Ideology and is Explained by Moral

- Judgment' (2023) 0 (0) Personality and Social Psychology Bulletin 1
<<https://doi.org/10.1177/01461672231198001>> accessed 15 October 2023
- Cater D, 'Toward a Public Philosophy on Government-Media Relations' in William L Rivers & Michael J Nyhan (eds), *Aspen Notebook on Government and the Media* (Praeger London 1973) 3
- Chaiken S and Eagly A H, 'Communication Modality as a Determinant of Persuasion: The Role of Communicator Saliency' (1983) 45 (2) J of Personality and Social Psychology 241
- Chastagner C, 'The Parents' Music Resource Center: From Information to Censorship' (May 1999) 18(2) Popular Music 179
- Cichocka A, 'Understanding defensive and secure in-group positivity: The role of collective narcissism' (2016) 27(1) European Review of Social Psychology 283
- CloudResearch (2021) <<https://www.cloudresearch.com>> last accessed 4 May 2021
- Cohen D, 'Senate report on Russian interference was written by disinformation warriors behind Alabama 'false flag operation' (The Grayzone, 25 December 2018)
<<https://thegrayzone.com/2018/12/25/senate-report-on-russian-interference-was-written-by-disinformation-warriors-behind-alabama-false-flag-operation/>> accessed 28 July 2022
- Collins T P, Crawford J T, and Brandt M J, 'No evidence for ideological asymmetry in dissonance avoidance: Unsuccessful close and conceptual replications of Nam, Jost, and Van Bavel' (2013) 48 (3) Social Psychology 123
- Conover M D and others, 'Political Polarization on Twitter' (2011) Proceedings of the Fifth International Association for the Advancement of Artificial Intelligence Conference on Weblogs and Social Media 89
- Cossé E, 'The Alarming Rise of Anti-Semitism in Europe' (Human Rights Watch, 04 June 2019) <<https://www.hrw.org/news/2019/06/04/alarming-rise-anti-semitism-europe>> accessed 29 October 2022
- Costarelli S, 'The Pros and Cons of Ingroup Ambivalence: The Moderating Roles of Attitudinal Basis and Individual Differences in Ingroup Attachment and Glorification' (2015) 23 Current Research in Social Psychology 26

- Costarelli S and Colis E, 'In-group Attachment and Glorification, Perceptions of Cognition-based ambivalence as Contributing to the Group, and Positive Affect' (2016) 24 Current Research in Social Psychology 31
- Craufurd Smith R and Tambini D, 'Measuring Media Plurality in the United Kingdom: Policy Choices and Regulatory Challenges' (2012) 4(1) Journal of Media Law 35
- Crawford K and Gillespie T, 'What is a flag for? Social media reporting tools and the vocabulary of complaint' (2016) 18(3) new media & society 410
- Crawford M T, 'The renegotiation of social identities in response to a threat to self-evaluation maintenance' (2007) 43 J Experimental Social Psychology 39
- Crawford S, 'First Amendment Common Sense' (30 September 2014) 127 Harvard L Rev 2343
- Croce M and Salvatore A, 'Why Does the Law Want Us to be Normal? Schmitt's Institutionalism and the Critique of the Liberal Legal Order' (Spring 2016) 93 Cultural Critique 32
- Crocker T P, 'Displacing Dissent: The Role of "Place" in First Amendment Jurisprudence' (2006-07) 75 Fordham L Rev 2587
- Crozier M, 'Western Europe' in *The Crisis Of Democracy* by Michel Crozier, Samuel P Huntington, & Joji Watanuki (1975 NYU Press) 11
- Curran J, 'Media and the Making of British Society, c. 1700-2000' (2002) 8(2) Media History 135
- Czitrom D J, *Media and the American Mind: From Morse to McLuhan* (Univ N Carolina P 2010)
- Democratic National Committee, 'The 2020 Democratic Platform' (2020)
<<https://democrats.org/where-we-stand/party-platform/>> last accessed 4 May 2021
- Denham E, 'Letter to Julian Knight MP, Chair, Digital, Culture and Media and Sport Select Committee' (Information Commissioner's Office, 02 October 2020)
<https://ico.org.uk/media/action-weve-taken/2618383/20201002_ico-o-ed-l-rtl-0181_to-julian-knight-mp.pdf> accessed 28 August 2022

Department for Digital, Culture, Media & Sport, 'Media Literacy Taskforce Fund' (18 July 2022) <<https://www.gov.uk/guidance/media-literacy-taskforce-fund>> accessed 28 July 2022

Department for Digital, Culture, Media & Sport and Dorries N, 'Tougher protections for journalism added to online safety laws' (07 July 2022) <<https://www.gov.uk/government/news/tougher-protections-for-journalism-added-to-online-safety-laws>> accessed 28 July 2022

Department for Digital, Culture, Media & Sport and Dorries N, 'Online Safety Update' Statement UIN HCWS194 (07 July 2022) <<https://questions-statements.parliament.uk/written-statements/detail/2022-07-07/hcws194>> accessed 28 August 2022

Department for Digital, Culture, Media & Sport and Lopez J, 'UK digital identity and attributes trust framework – beta version' (13 June 2022) <<https://www.gov.uk/government/publications/uk-digital-identity-and-attributes-trust-framework-beta-version>> accessed 28 August 2022

Department for Digital, Culture, Media & Sport and Dorries N, 'World-first online safety laws introduced in Parliament' (17 March 2022) <<https://www.gov.uk/government/news/world-first-online-safety-laws-introduced-in-parliament>> accessed 28 August 2022

Department for Digital, Culture, Media & Sport and Warman M, 'Government sets out new plans to help build trust in use of digital identities' (11 February 2021) <<https://www.gov.uk/government/news/government-sets-out-new-plans-to-help-build-trust-in-use-of-digital-identities>> accessed 28 July 2022

Department for Digital, Culture, Media and Sport, *Online Harms* (White Paper, CP 57, April 2019)

Devine M, 'Media helped hide the real Joe Biden by censoring Hunter stories: Devine' (New York Post, 28 November 2021) <<https://nypost.com/2021/11/28/media-helped-hide-the-real-joe-biden-by-censoring-hunter-stories-devine/>> accessed 01 November 2022

- Dhir A and others, 'Why do people share fake news? Associations between the dark side of social media use and fake news sharing behavior' (2019) 51 J Retailing and Consumer Services 72
- Dixon F H, 'The Mann-Elkins Act, Amending the Act to Regulate Commerce' (August 1910) 24(4) Quarterly Journal of Economics 593
- Directorate-General for Communications Networks, Content and Technology, 'Shaping Europe's digital future: 2018 Code of Practice on Disinformation' (21 July 2022) <<https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation>> accessed 28 August 2022
- Doctorow C, 'Right or Left, You Should Be Worried About Big Tech Censorship' (*Electronic Frontier Foundation*, 16 July 2021) <<https://www.eff.org/deeplinks/2021/07/right-or-left-you-should-be-worried-about-big-tech-censorship>> accessed 28 July 2022
- Dover T L and others, 'Does unfairness feel different if it can be linked to group membership? Cognitive, affective, behavioral and physiological implications of discrimination and unfairness' (2014) 56 J Experimental Psychology 96
- Drale C S, 'The United Fruit Company and Early Radio Development' (2010) 17(2) J Radio & Audio Media 195
- Duarte N, Llanso E, and Loup A, 'Mixed Messages? The Limits of Automated Social Media Content Analysis' (2018) 81 Proceedings of Machine Learning Research, 1st Conference on Fairness, Accountability and Transparency
- Durant W, *The Story of Civilization, Vol. IV: The Age of Faith* (Simon & Schuster 1950)
- Edelman R, 'Fake News; the Neutron Bomb Explodes' (09 November 2017) <<https://www.edelman.com/insights/fake-news-neutron-bomb-explodes>> accessed 28 July 2022
- Edelman, '2021 Edelman Trust Barometer UK Findings' (Edelman, 22 February 2021) <<https://www.edelman.co.uk/edelman-trust-barometer-2021-uk-findings>>
- Effron D A and Raj M, 'Misinformation and Morality: Encountering Fake-News Headlines Makes Them Seem Less Unethical to Publish and Share' (2020) 31 (1) Psychological Science 75

Eker I, Cichocka A, and Sibley C G, 'Investigating motivations underlying collective narcissism and in-group identification' (2023) 26(3) Group Processes & Intergroup Relations 669

Ellul J, Propaganda: The Formation of Men's Attitudes (Konrad Kellen and Jean Lerner trs, Vintage 1973)

England C, 'Did BBC Whistleblowers Mislead the Public on Labour Antisemitism and Blame Their Own Failures on Corbyn?' (Novara Media, 19 May 2020)
<<https://novaramedia.com/2020/05/19/did-bbc-whistleblowers-mislead-the-public-on-labour-antisemitism-and-blame-their-own-failures-on-corbyn/>> accessed 01 November 2022

Equality and Human Rights Commission, 'Investigation into antisemitism in the Labour Party' (October 2020)

Facebook, 'Facebook Reports Fourth Quarter and Full Year 2020 Results' (27 January 2021)
<<https://investor.fb.com/investor-news/press-release-details/2021/Facebook-Reports-Fourth-Quarter-and-Full-Year-2020-Results/default.aspx>> accessed 28 July 2022

Facebook, 'Taking Action Against People Who Repeatedly Share Misinformation' (26 May 2021)
<<https://about.fb.com/news/2021/05/taking-action-against-people-who-repeatedly-share-misinformation/>> accessed 28 July 2022

Farhall K and others, 'Political Elites' Use of Fake News Discourse Across Communications Platforms' (2019) 13 Int'l J of Comms 4353

Federal Communications Commission, 'History of Commercial Radio' (April 12, 2021)
<<https://www.fcc.gov/media/radio/history-of-commercial-radio>> accessed 05 July 2022

Federal Communications Commission, 'FCC Chairman Genachowski continues regulatory reform to ease burden on businesses; announces elimination of 83 outdated rules' (22 August 2011) DOC 309224A1 <<https://docs.fcc.gov/public/attachments/DOC-309224A1.pdf>> accessed 05 July 2022

Federal Communications Commission, 'Program Content Regulations' (29 September 2017)
<<https://www.fcc.gov/media/program-content-regulations>> accessed 05 July 2022

Federal Trade Commission, 'FTC Imposes \$5 Billion Penalty and Sweeping New Privacy Restrictions on Facebook' (24 July 2019) <<https://www.ftc.gov/news-events/news/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions-facebook>> accessed 28 August 2022

Federal Trade Commission, 'FTC Sues Cambridge Analytica, Settles with Former CEO and App Developer' (24 July 2019) <<https://www.ftc.gov/news-events/news/press-releases/2019/07/ftc-sues-cambridge-analytica-settles-former-ceo-app-developer>> accessed 28 August 2022

Fernandez J M, 'Malaysia's Anti-Fake News Act' (2019) 25 *Pacific Journalism Review* 173

Fisher M, 'Inside Facebook's Secret Rulebook for Global Political Speech' (*New York Times*, 27 December 2018) <<https://www.nytimes.com/2018/12/27/world/facebook-moderators.html>> accessed 28 July 2022

Fisher R, 'Press Rights vs. Press Responsibilities: The Issues of the First Amendment' in William L. Rivers & Michael J. Nyhan (eds), *Aspen Notebook on Government and the Media* (Praeger London 1973) 51

Fisher R, 'The Public's Right to Know: Functions and Malfunctions of the Media' in William L. Rivers & Michael J. Nyhan (eds), *Aspen Notebook on Government and the Media* (Praeger London 1973) 78

Fischer S, 'Facebook says new pre-election political ad rules apply to boosted posts' (*Axios*, 27 October 2020) <<https://www.axios.com/2020/10/27/facebook-says-new-pre-election-political-ad-rules-apply-to-boosted-posts>> accessed 28 July 2022

Fischer S, 'Technical glitch in Facebook's ad tools creates political firestorm' (*Axios*, 30 October 2020) <<https://www.axios.com/2020/10/30/technical-glitch-facebooks-ad-tool-political-election>> accessed 28 July 2022

Fischer S, "'Unreliable" news sources got more traction in 2020' (*Axios*, 22 December 2020) <<https://www.axios.com/2020/12/22/unreliable-news-sources-social-media-engagement>> accessed 28 July 2022

Fletcher R and others, 'Measuring the reach of "fake news" and online disinformation in Europe' Reuters Institute for the Study of Journalism at the University of Oxford (February 2018)

Forde Inquiry Panel, *The Forde Report* (19 July 2022) <<https://www.fordeinquiry.org/forde-inquiry-report/>> accessed 01 November 2022

Foreign, Commonwealth & Development Office, 'Counter Disinformation and Media Development Programme' (2021)
<[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/842428/Counter Disinformation and Media Development Programme.odt](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/842428/Counter_Disinformation_and_Media_Development_Programme.odt)>

Freierman S, 'Telegram Falls Silent Stop Era Ends Stop' (*New York Times*, 06 Feb 2006)
<<https://www.nytimes.com/2006/02/06/technology/telegram-falls-silent-stop-era-ends-stop.html>> accessed 01 April 2023

Frenkel S, Casey N, and Mozur P, 'In Some Countries, Facebook's Fiddling Has Magnified Fake News' (*New York Times*, 14 January 2018)
<<https://www.nytimes.com/2018/01/14/technology/facebook-news-feed-changes.html>> accessed 28 July 2022

Frerichs S, 'Putting behavioural economics in its place: the new realism of law, economics and psychology and its alternatives' (2021) 72(4) NILQ 651

Gabicka A and Heinemann A, 'Social Media Competition Law' (2014) 37(2) World Competition L and Econ Rev 149

Gallup, 'Party Affiliation' (2022) <<https://news.gallup.com/poll/15370/party-affiliation.aspx>> accessed 28 July 2022

Gardiner H, *Mysteries' End: An Investigation of the Last Days of the Medieval Religious Stage* (Yale 1946)

Gardner A, 'Trump pressured a Georgia elections investigator in a separate call legal experts say could amount to obstruction' (*Washington Post*, 11 March 2021)
<https://www.washingtonpost.com/politics/trump-call-georgia-investigator/2021/01/09/7a55c7fa-51cf-11eb-83e3-322644d82356_story.html>
accessed 28 July 2022

Garnett H, 'Presentation' (Politics on Wednesdays, UEA School of Politics, Philosophy, and Language, 28 September 2022)

- Gaughan A J, 'Illiberal Democracy: The Toxic Mix of Fake News, Hyperpolarization, and Partisan Election Administration' (2017) 12 Duke J Const L & Pub Pol'y 59
- Geller H and Lampert D, 'Cable, Content Regulation and the First Amendment' (1983) 32(3) Catholic Univ L Rev 603
- Gil de Zuniga H, Jung N, Valenzuela S, 'Social Media Use for News and Individuals' Social Capital, Civic Engagement and Political Participation' (2012) 17 J Computer-Mediated Communication 319
- Gillespie T, 'Governance of and by platforms' in Jean Burgess, Alice Marwick, and Thomas Poell (eds), *The SAGE Handbook of Social Media* (SAGE 2018)
- Gilmore K, 'Must We Have UHF-TV?' (May 1962) Popular Electronics, *available at* <https://www.rfcafe.com/references/popular-electronics/must-we-have-uhf-tv-may-1962-popular-electronics.htm> accessed 05 July 2022
- Goldmacher S and Conger K, 'Under G.O.P. Pressure, Tech Giants Are Empowered by Election Agency' (New York Times, 13 September 2021) <https://www.nytimes.com/2021/09/13/us/politics/fec-twitter-hunter-biden.html> accessed 01 November 2022
- Google, 'Removals under the Network Enforcement Law' (*Google.com*, 31 December 2019) https://transparencyreport.google.com/netzdg/youtube?hl=en_GB accessed 28 July 2022
- Google Trends, 'Interest over time: labour antisemitism, 01 October 2007 – 01 November 2022' (Google, 2022) <https://trends.google.com/trends/explore?date=2007-10-01%202022-11-01&geo=GB&q=labour%20antisemitism> accessed 01 November 2022
- Gordillo F and Mestas L, 'What we know about people shapes the inferences we make about their personalities' (2015) 23 Current Research in Social Psychology 38
- Graeber D, *The Utopia of Rules: On Technology, Stupidity, and the Secret Joys of Bureaucracy* (Melville House 2015)
- Graeber D, 'The Center Blows Itself Up: Care and Spite in the 'Brexit Election' (New York Review of Books, 13 January 2020)

<<https://www.nybooks.com/daily/2020/01/13/the-center-blows-itself-up-care-and-spite-in-the-brexitelection/>> accessed 28 July 2022

Grant F and Hogg M A, 'Self-uncertainty, social identity prominence and group identification' (2012) 38 J Experimental Social Psychology 538

Griffin A, 'Facebook PR boss Nick Clegg defends its decision to allow false political ads' (Independent, 20 January 2020) <<https://www.independent.co.uk/life-style/gadgets-and-tech/facebook-nick-clegg-political-ads-2020-election-a9292891.html>> accessed 28 July 2022

Griffin A, 'Facebook puts groups on "probation" as it tries to stop dangerous content in wake of election results' (*Independent*, 09 November 2020) <<https://www.independent.co.uk/tech/facebook-groups-us-election-probation-b1719839.html>> accessed 28 July 2022

Grinberg N and others, 'Fake News on Twitter during the 2016 U.S. presidential election' (2019) 363 Science 374

Guardian staff and agencies, "'It's the real me": Nigerian president denies dying and being replaced by clone' (The Guardian, 03 December 2018) <<https://www.theguardian.com/world/2018/dec/03/its-real-me-nigerian-president-denies-dying-and-being-replaced-by-clone>> accessed 28 July 2022

Guess A, Nagler J, and Tucker J, 'Less than you think: Prevalence and predictors of fake news dissemination on Facebook' (2019) 5(1) Science Advances <<https://www.science.org/doi/10.1126/sciadv.aau4586>> accessed 28 July 2022

Guess A, Nyhan B, and Reifler J, 'Selective Exposure to Misinformation: Evidence from the consumption of fake news during the 2016 U.S. presidential campaign' (9 January 2018) 1 <<https://www.dartmouth.edu/~nyhan/fake-news-2016.pdf>> accessed 02 June 2020

Grunes A P and Stucke M E, 'No Mistake About It: The Important Role of Antitrust in the Era of Big Data' (*Antitrust Source*, April 2015) University of Tennessee Legal Studies Research Paper No 269

Guzman G G, 'Household Income: 2018' (U.S. Census Bureau, Sept. 2019) American Community Survey Briefs ACSBR/18-01

- Habermas J, "The Public Sphere: An Encyclopedia Article (1964)" (1974) 3 New German Critique 49
- Hackel L M and others, 'From groups to grits: Social identity shapes evaluations of food pleasantness' (2018) 74 J Experimental Social Psychology 270
- Hadfield G K, *Rules for a Flat World* (OUP 2016)
- Lady Justice Hallett, 'Trial by Jury – Past and Present' (35th Blackstone Lecture, 22 May 2017) <<https://www.judiciary.uk/wp-content/uploads/2017/05/hallett-lj-blackstone-lecture-20170522-1.pdf>> accessed 31 May 2020
- Hallin D C & Mancini P, *Comparing Media Systems: Three Models of Media and Politics* (Cambridge UP 2004)
- Haney C, 'Psychology and Legal Change: On the Limits of a Factual Jurisprudence' (1980) 4(3) Law and Human Behaviour 147
- Hardy J, *Western Media Systems* (Routledge 2008)
- Harris B J, 'Content Rated by: An Oral History of the ESRB' (2019-2020) <<https://www.esrb.org/about/>> accessed 05 July 2022
- Harvard Law Review Ass'n, 'The Freedom of Radio Speech,' (1933) 46(6) Harvard L Rev 987-993
- Hayek F A, 'The Use of Knowledge in Society' (Sept. 1945) 35(4) American Economic Review 519
- Hazard Owen L, 'Facebook drastically changes News Feed to make it "good for people" (and bad for most publishers)' (*NiemanLab*, 11 January 2018) <<https://www.niemanlab.org/2018/01/facebook-drastically-changes-news-feed-to-make-it-good-for-people-and-bad-for-most-publishers/>> accessed 28 July 2022
- Hazard Owen L, 'As traditional news use declines, online news isn't making up the gap' (*NiemanLab*, 14 June 2022) <<https://www.niemanlab.org/2022/06/as-traditional-news-use-declines-online-news-isnt-making-up-the-gap/>> accessed 28 July 2022
- Hazlett T, *The Political Spectrum: The Tumultuous Liberation of Wireless Technology, from Herbert Hoover to the Smartphone* (Yale UP 2017)

- Hearst D, 'Geoffrey Bindman on Labour's antisemitism scandal: "There has been political manipulation"' (Middle East Eye, 26 August 2022)
<<https://www.middleeasteye.net/news/labour-antisemitism-geoffrey-bindman-political-manipulation>> accessed 01 November 2022
- Helm R K and Nasu H, 'Regulatory Responses to "Fake News" and Freedom of Expression: Normative and Empirical Evaluation' (2021) 21 Human Rights Law Review 302
- Hermann E S and Chomsky N, *Manufacturing Consent* (Pantheon Books, updated ed. 2002)
- Hildreth J A D and Anderson C, 'Does loyalty trump honesty? Moral judgements of loyalty-driven deceit' (2018) 79 J Experimental Social Psychology 87
- Hindman M, *The Internet Trap: How the Digital Economy Builds Monopolies and Undermines Democracy* (Princeton UP 2018)
- Hintz A, 'Restricting digital sites of dissent: commercial social media and free expression' (2016) 13 Critical Discourse Studies 325
- Hogg M A, 'Subjective Uncertainty Reduction through Self-categorization: A Motivational Theory of Social Identity Processes' Processes' in Wolfgang Stroebe and Miles Hewstone (eds), 11 *European Review of Social Psychology* (Wiley 2000) 223
- Holmes S and Ganley P, 'User-generated content and the law' (2007) 2(5) J Intellectual Property Law & Practice 338
- Home Affairs Committee, 'Antisemitism in the UK' HC 136 (13 October 2016)
- Homeland Security Advisory Council, Disinformation Best Practices and Safeguards Subcommittee, *Final Report* (2022) <https://www.dhs.gov/sites/default/files/2022-08/22_0824_ope_hsic-disinformation-subcommittee-final-report-08242022.pdf> accessed 15 November 2022
- Horne B D, Norregaard J, and Adali S, 'Robust Fake News Detection Over Time and Attack' (2019) 11(1) Association for Computing Machinery Transactions on Intelligent Systems and Technology, article 7
- Horwitz J and Seetharaman D, 'Facebook Executives Shut Down Efforts to Make the Site Less Divisive' (*Wall Street Journal*, 26 May 2020)

<<https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499>> accessed 28 July 2022

Hughes C, 'Opinion: It's Time to Break Up Facebook' (*New York Times*, 09 May 2019) <<https://www.nytimes.com/2019/05/09/opinion/sunday/chris-hughes-facebook-zuckerberg.html>> accessed 28 July 2022

Huntington S P, 'The United States' in *The Crisis Of Democracy* by Michel Crozier, Samuel P Huntington, & Joji Watanuki (1975 NYU Press) 59

Huq A and Ginsberg T, 'How to Lose a Constitutional Democracy' (2018) 65 UCLA Law Review 78

Hutchison P and others, 'Getting rid of the bad ones: The relationship between group identification, deviant derogation, and identity maintenance' (2008) 44 J Experimental Social Psychology 874

Ice T, 'Status' (23 February 2019) <<https://x.com/FINALLEVEL/status/1099389177483878401>> accessed 30 November 2023

Jack C, 'Lexicon of Lies: Terms for Problematic Information' (Data & Society Research Institute, August 2017) 1 <<https://datasociety.net/library/lexicon-of-lies/>> accessed 28 July 2022

Jagodzinski C, *Privacy and Print: Reading and Writing in Seventeenth Century England* (Virginia UP 1999)

Jans L and others, 'Dynamic Reactions to opinion deviance: The role of social identity formation' (2019) 84 J Experimental Social Psychology 103803

Jefferson T, 'Letter to John Norvell' (11 June 1807) Founders Online, National Archives <<https://founders.archives.gov/documents/Jefferson/99-01-02-5737>> accessed 28 July 2022

Jenkins Jr. H W, 'Media Bias and Hunter's Laptop' (*Wall Street Journal*, 12 April 2022) <<https://www.wsj.com/articles/media-spin-hunter-biden-laptop-news-trump-free-speech-censorship-stop-the-steal-big-lie-2020-presidential-election-steele-dossier-11649792758>> accessed 01 November 2022

- The Joe Rogan Experience*, '#1863 – Mark Zuckerberg' (Spotify, 25 August 2022) <<https://open.spotify.com/episode/51gxrAActH18RGhKNza598?si=a61ababed4584e76>> accessed 28 August 2022
- Jost J T, 'Ideological Asymmetries and the Essence of Political Psychology' (2017) 38 (2) 167
- Jost J T and others, 'Political Conservatism as Motivated Social Cognition' (2003) 129 (3) Psychological Bulletin 339
- Jost J T, Nosek B A, and Gosling S D, 'Ideology: Its Resurgence in Social, Personality, and Political Psychology' (2008) 3 (2) Perspectives on Psychological Science 126
- Juris J S, 'Reflections on #Occupy Everywhere: Social media, public space, and emerging logics of aggregation' (2012) 39(2) American Ethnologist 259
- Kaczmarek L and others, 'Social Media Monitoring of the Campaigns for the 2013 German Bundestag Elections on Facebook and Twitter' (Leibniz Institute for the Social Sciences, 31 March 2014) GESIS Working Papers 31
- Kahan D M, 'Ideology, motivated reasoning, and cognitive reflection' (2013) 8 (4) Judgment and Decision Making 407
- Kan M, 'Twitter: Our "Read Before You Retweet" Function Actually Works' (*PCMag.com*, 25 September 2020) <<https://www.pcmag.com/news/twitter-our-read-before-you-retweet-function-actually-works>> last accessed 28 April 2021
- Kang J and others, 'Implicit Bias in the Courtroom' (2012) 59 UCLA L Rev 1124
- Kielbowicz R B, 'News Gathering by Mail in the Age of the Telegraph: Adapting to a New Technology' (January 1987) 28(1) Technology and Culture 26
- Kim A and Dennis A R, 'Says Who? The Effects of Presentation Format and Source Rating on Fake News in Social Media' (2019) 43(3) Management Information Systems Quarterly 1025
- Klemfuss J Z and others, 'The Effects of Exposure to Differing Amounts of Misinformation and Source Credibility Perception on Source Monitoring and Memory Accuracy' (2017) 4 (4) Psychology of Consciousness: Theory, Research, and Practice 337
- Klonick K, 'The New Governors: The People, Rules, and Processes Governing Online Speech' (2018) 131 Harv L Rev 1598

Knowles E, *What They Didn't Say: A Book of Misquotations* (Oxford UP 2006)

Körner A, Joffe S, and Deutsch R, 'When skeptical, stick with the norm: Low dilemma plausibility increases deontological moral judgments' (2019) 84 J Experimental Social Psychology 103834

Kroll J A and others, 'Accountable Algorithms' (2017) 165 U Pennsylvania L Rev 633

Kyprianides A, Easterbrook M J, and Brown R, 'Group identities benefit well-being by satisfying needs' (2019) 84 J Experimental Social Psychology 103836

Lacy D, 'Rules for the Regulators: A Rationale for Government's Role' in William L Rivers & Michael J Nyhan (eds), *Aspen Notebook on Government and the Media* (Praeger London 1973) 93

Lantian A and others, 'Measuring Belief in Conspiracy Theories: Validation of a French and English Single-Item Scale' (2016) 29(1) *International Review of Social Psychology* 1

Lassetter B and Neel R, 'Malleable liberals and fixed conservatives? Political orientation shapes perceived ability to change' (2019) 82 J Experimental Social Psychology 141

Lavi M, 'Content Providers' Secondary Liability: A Social Network Perspective' (2016) 26(4) *Fordham Intellectual Property, Media & Entertainment Law Journal* 855

Lessig L, *Code: Version 2.0* (Basic Books/Perseus 2006)

Lewins W, *Her Majesty's Mails: A History of the Post-Office and an Industrial Account of its Present Condition* (2nd edn, Sampson Low 1865)

Lichfield G, 'EmTech Stage: Twitter's CTO on misinformation' (*MIT Technology Review*, 18 November 2020)
<<https://www.technologyreview.com/2020/11/18/1012066/emtech-stage-twitters-cto-on-misinformation/>> accessed 28 July 2022

Likert R, 'A Technique for the Measurement of Attitudes' in R S Woodworth (ed) *Archives of Psychology* 22 (1932-33) 5

Lindholm T and Christianson S-A, 'Intergroup Biases and Eyewitness Testimony' (1998) 136(6) J Social Psychology 710

Lippmann W, *Public Opinion* (Simon & Schuster 1997)

- Lipton E and Broder J M, 'Regulator Deferred to Oil Industry on Rig Safety' (*New York Times*, 7 May 2010) <<https://www.nytimes.com/2010/05/08/us/08agency.html>> accessed 25 April 2023
- Liu J, Thomas J M, and Higgs S, 'The relationship between social identity, descriptive social norms and eating intentions and behaviors' (2019) 82 J Experimental Social Psychology 217
- Loftus E F, 'Eyewitness testimony' (2019) 33 Applied Cognitive Psychology 498
- Lord C G, Ross L, and Lepper M R, 'Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence' (1979) 37 (11) J of Personality and Social Psychology 2098
- Louw E, *The Media and Cultural Production* (SAGE 2001)
- Luchins A S and Luchins E H, 'Social influences on judgment of changing evidence' (1961) 54 J Social Psychology 13
- Mack Z, 'Why AI can't fix content moderation' (*The Verge*, 02 July 2019) <<https://www.theverge.com/2019/7/2/20679102/content-moderation-ai-social-media-behind-the-screen-sarah-t-roberts-vergecast>> accessed 28 July 2022
- Mackie D M, Smith E R, and Ray D G, 'Intergroup Emotions and Intergroup Relations' (2008) 2/5 Social and Personality Psychology Compass 1866
- MacKinnon R, 'Flatter world and thicker walls? Blogs, censorship and civic discourse in China' (2008) 134 Public Choice 31
- Mackintosh E, 'Analysis: No matter who wins the US election, the world's "fake news" problem is here to stay' (CNN, 25 October 2020) <<https://edition.cnn.com/2020/10/25/world/trump-fake-news-legacy-intl/index.html>> accessed 28 July 2022
- MacLeod A, 'Meet the ex-CIA Agents Deciding Facebook's Content Policy' (*MintPress News*, 12 July 2022) <<https://www.mintpressnews.com/meet-ex-cia-agents-deciding-facebook-content-policy/281307/>> accessed 01 April 2023
- Mac Síthigh D, *Medium Law* (Routledge 2018)

- Madison J, 'The Same Subject Continued: The Union as a Safeguard Against Domestic Faction and Insurrection' *The Federalist Papers No. 10* (23 November 1787)
- Maitner A T and others, 'Identity salience moderates processing of group-relevant information' (2010) 46 *J Experimental Social Psychology* 441
- Manchester C, 'Lord Campbell's Act: England's First Obscenity Statute' (1988) 9(2) *J of Legal History* 223
- Manskar N, 'Twitter, Facebook censor Post over Hunter Biden expose' (New York Post, 14 October 2020) <<https://nypost.com/2020/10/14/facebook-twitter-block-the-post-from-posting/>> accessed 01 November 2022
- Manskar N, 'Jack Dorsey says blocking Post's Hunter Biden story was "total mistake" – but won't say who made it' (New York Post, 25 March 2021) <<https://nypost.com/2021/03/25/dorsey-says-blocking-posts-hunter-biden-story-was-total-mistake/>> accessed 28 July 2022
- Marcellino W and others, 'Detecting Conspiracy Theories on Social Media' (RAND Corporation, 2021) <<https://www.rand.org/t/RAA676-1>> accessed 28 July 2022
- Maté A, 'New Studies Show Pundits Are Wrong About Russian Social-Media Involvement in US Politics' (*The Nation*, 28 December 2018) <<https://www.thenation.com/article/archive/russiagate-elections-interference/>> accessed 28 July 2022
- McCright A M & Dunlap R E, 'Combating Misinformation Requires Recognizing Its Types and the Factors That Facilitate Its Spread and Resonance' (2017) 6 *J of Applied Research in Memory and Cognition* 389
- McCullagh K, 'E-Democracy: Potential for Political Revolution?' (2003) 11(2) *International Journal of Law and Information Technology* 149
- McGuire W J, 'Resistance to Persuasion Conferred by Active and Passive Prior Refutation of the Same and Alternative Counterarguments' (1961) 63 (2) *J of Abnormal and Social Psychology* 326
- McLuhan M, *Letters of Marshall McLuhan* (Oxford UP 1987)

- McNamee R, 'Platforms Must Pay for Their Role in the Insurrection' (*Wired*, 07 January 2021) <<https://www.wired.com/story/opinion-platforms-must-pay-for-their-role-in-the-insurrection/>> accessed 28 July 2022
- McNicholas A and Ward D, 'United Kingdom: New Media' in Bernd-Peter Lange and David Ward (eds), *The Media and Elections: A Handbook and Comparative Study* (Taylor and Francis 2004) 159
- McQuail D, *Media Accountability and Freedom of Publication* (OUP 2007)
- McSherry C, York J C, and Cohn C, 'Private Censorship is Not the Best Way to Fight Hate or Defend Democracy: Here Are Some Better Ideas' (*Electronic Frontier Foundation*, 30 January 2018) <<https://www.eff.org/deeplinks/2018/01/private-censorship-not-best-way-fight-hate-or-defend-democracy-here-are-some>> accessed 26 July 2022
- Meta, 'Information on prohibited ads related to voting and ads about social issues, elections or politics' (*Meta Business Help Centre* 2022) <<https://www.facebook.com/business/help/253606115684173>> accessed 28 August 2022
- Mills A J, Pitt C, and Lord Ferguson S, 'The Relationship between Fake News and Advertising: Brand Management in the Era of Programmatic Advertising and Prolific Falsehood' (March 2019) *Journal of Advertising Research* 3
- Minkkinen P, 'Agonism, Democracy, and Law' in Simon Stern and others, eds *The Oxford Handbook of Law and Humanities* (2019) 426
- Mitchell L M, 'Who Watches the Watchdog?: The Crisis of Public Confidence' in William L Rivers & Michael J Nyhan (eds), *Aspen Notebook on Government and the Media* (Praeger London 1973) 144
- Moore M, 'YouTube deletes Iranian channel Press TV UK for flouting ban' (*The Times*, 29 January 2020) <<https://www.thetimes.co.uk/article/youtube-deletes-iranian-channel-press-tv-uk-for-flouting-ban-gxz539v80>> accessed 28 August 2022
- Morris E and Fonrouge G, 'Smoking-gun email reveals how Hunter Biden introduced Ukrainian businessman to VP dad' (New York Post, 14 October 2020) <<https://nypost.com/2020/10/14/email-reveals-how-hunter-biden-introduced-ukrainian-biz-man-to-dad/>> accessed 01 November 2022

- Morris R C, 'Identity Salience and Identity Importance in Identity Theory' (2013) 21 Current Research in Social Psychology 23
- Mosca G, *The Ruling Class* (McGraw-Hill 1939)
- Mumford L, *Technics and Civilization* (7th edn, Routledge & Kegan Paul 1955)
- Munger K, 'Social Media, Political Science, and Democracy' (2019) 81(3) Journal of Politics 62
- Murphy G and others, 'False Memories for Fake News During Ireland's Abortion Referendum' (2019) 30 (10) Psychological Science 1449
- Murray A, *Information Technology Law: The Law & Society* (2nd edn, OUP 2013)
- Musk E, 'Status' (01 November 2022)
<<https://x.com/elonmusk/status/1587498907336118274>> accessed 30 November 2023
- Naar I, 'Saudi Arabia's government tops Edelman's 2021 Trust Barometer report' (Al Arabiya, 27 March 2021) <<https://english.alarabiya.net/News/gulf/2021/03/27/Saudi-Vision-2030-Saudi-Arabia-s-government-tops-Edelman-s-2021-Trust-Barometer-report>> accessed 28 July 2022
- Napoli P M, 'Lessons for social media from the Fairness Doctrine,' (*Columbia Journalism Review*, 13 August 2020) <<https://www.cjr.org/opinion/social-media-fairness-doctrine.php>> accessed 05 July 2022
- Napoli P M, 'What Would Facebook Regulation Look Like? Start With the FCC' (*Wired*, 04 October 2019) <<https://www.wired.com/story/what-would-facebook-regulation-look-like-start-with-the-fcc/>> accessed 26 September 2022
- Napoli P M, *Social Media and the Public Interest: Media Regulation in the Disinformation Age* (2019, Columbia UP)
- National Intelligence Council, 'Foreign Threats to the 2020 US Federal Elections' (10 March 2021) Intelligence Community Assessment 2020-00078D
<<https://www.dni.gov/files/ODNI/documents/assessments/ICA-declass-16MAR21.pdf>> accessed 28 July 2022
- Nederhof A J, 'Methods of Coping with Social Desirability Bias: A Review' (1985) 15 Eur J Social Psychology 263

The New York Times, 'Worldwide Propaganda Network Built by the C.I.A.' (*New York Times*, 26 December 1977) <<https://www.nytimes.com/1977/12/26/archives/worldwide-propaganda-network-built-by-the-cia-a-worldwide-network.html>> accessed 28 July 2022

Nichol A, Millar G, and Sharland A, *Media Law and Human Rights* (2nd edn, OUP 2009)

Nichols R M and Loftus E F, 'Who is susceptible in three false memory tasks?' (2019) 27 (7) *Memory* 962

Nokia Bell Labs, 'Telstar 1' (2023) <<https://www.bell-labs.com/about/history/innovation-stories/telstar-1/>> accessed 25 March 2023

Norden E and McLuhan M, 'Playboy Interview: Marshall McLuhan—A Candid Conversation with the High Priest of Popcult and Metaphysician of Media' (1969) 16(3) *Playboy* 53

Nyhan B, 'Why Fears of Fake News Are Overhyped' (GEN, 04 February 2019) <<https://gen.medium.com/why-fears-of-fake-news-are-overhyped-2ed9ca0a52c9>> accessed 28 July 2022

Nyhan B, 'We Can't Combat Fake News If We Don't Really Understand It' (GEN, 07 March 2019) <<https://gen.medium.com/we-cant-combat-fake-news-if-we-don-t-really-understand-it-caed33dbea09>> accessed 28 July 2022

Obar J and Oeldorf-Hirsch A, 'The Clickwrap: A Political Economic Mechanism for Manufacturing Consent on Social Media' (July–Sept 2018) *Social Media + Society* 1

O'Carroll L & Elgot J, 'Labour pays out six-figure sum and apologises in antisemitism row' (*The Guardian*, 22 July 2020) <<https://www.theguardian.com/politics/2020/jul/22/labour-pays-out-six-figure-sum-and-apologises-in-antisemitism-row>> accessed 01 November 2022

Ofcom, *Adults' Media Use and Attitudes Report 2018* (25 April 2018)

Oliver J and Gilby N, 'Revealed: the secret British plan to keep Italy's communists from power' (*The Guardian*, 02 October 2022) <<https://www.theguardian.com/world/2022/oct/02/revealed-the-secret-british-plan-to-keep-italys-communists-from-power>> accessed 29 October 2022

Open Rights Group, 'ORG Policy Responses to Online Harms White Paper' (23 May 2019) <<https://www.openrightsgroup.org/publications/org-policy-responses-to-online-harms-white-paper/>> accessed 28 July 2022

Orbison C, "'Fighting Bob" Shuler: Early Radio Crusader' (1977) 21(4) J Broadcasting 459

Pagliaro S and others, 'Group based resiliency: Contrasting the negative effects of threat to the ingroup' (2013) 21 Current Research in Social Psychology 35

Palfrey J, 'Four Phases of Internet Regulation' (2010) in 77(3) *social research* 981

Paterson P, 'The Truth About Tonkin' (February 2008) 22(1) Naval History Magazine <<https://www.usni.org/magazines/naval-history-magazine/2008/february/truth-about-tonkin>> accessed 28 July 2022

Paul K, 'Facebook and Twitter restrict controversial New York Post story on Joe Biden' (*The Guardian*, 15 October 2020) <<https://www.theguardian.com/technology/2020/oct/14/facebook-twitter-new-york-post-hunter-biden>> accessed 28 July 2022

Paul K, 'Twitter and Facebook CEOs testify on alleged anti-conservative bias' (*The Guardian*, 17 November 2020) <<https://www.theguardian.com/us-news/2020/nov/17/twitter-facebook-senate-hearing-anti-conservative-bias>> accessed 28 July 2022

Paulhus D L and Vazire S, 'The Self-Report Method' in Richard W Robbins, R Chris Fraley, and Robert F Krueger (eds), *Handbook of Research Methods in Personality Psychology* (Guilford Press 2007) 224

Perraudin F, 'Twitter accuses Tories of misleading public with "factcheck" foray' (*The Guardian*, 20 November 2019) <<https://www.theguardian.com/politics/2019/nov/20/twitter-accuses-tories-of-misleading-public-in-factcheck-row>> accessed 29 October 2022

Pew Research Center, 'Political Independents: Who They Are, What They Think' (14 March 2019) <<https://www.pewresearch.org/politics/2019/03/14/political-independents-who-they-are-what-they-think/>> accessed 28 April 2021

Pew Research Center, 'Large differences in news sources trusted by Democrats and Republicans' (17 January 2020) <<https://www.pewresearch.org/journalism/2020/01/24/democrats-report-much-higher->

[levels-of-trust-in-a-number-of-news-sources-than-republicans/pj_2020-01-24_media-polarization_1-01/>](#) accessed 28 July 2022

Peterson E and Iyengar S, 'Partisan Gaps in Political Information and Information-Seeking Behavior: Motivated Reasoning or Cheerleading?' (2021) 65 (1) *Am J Pol Sci* 133

Petrocik J R, 'Measuring party support: Leaners are not independents' (2009) 28(4) *Electoral Studies* 562

Pickard V, 'The Fairness Doctrine won't solve our problems – but it can foster needed debate' (*The Washington Post*, 04 February 2021)
<<https://www.washingtonpost.com/outlook/2021/02/04/fairness-doctrine-wont-solve-our-problems-it-can-foster-needed-debate/>> accessed 05 July 2022

Pickard V, *Democracy without Journalism? Confronting the Misinformation Society* (2019 Oxford UP)

Pool I, 'Newsmen and Statesmen: Adversaries or Cronies?' in William L Rivers & Michael J Nyhan (eds), *Aspen Notebook on Government and the Media* (Praeger London 1973) 12

Postmes T, Haslam S A, and Jans L, 'A single-item measure of social identification: Reliability, validity, and utility' (2013) 52 *British J Social Psychology* 597

Press TV, 'Google again blocks access to YouTube account of Iran's Press TV' (19 September 2020) <<https://www.presstv.ir/Detail/2020/09/19/634536/Iran-Press-TV-YouTube-Google>> (accessed 28 August 2022)

Prime Research Solutions, 'How Does CloudResearch know Panel Demographic Data?' (CloudResearch.com, 2020) <<https://go.cloudresearch.com/knowledge/how-does-turkprime-know-panel-demographic-data>> last accessed 4 May 2021

Qualtrics XM (2021) <<https://www.qualtrics.com/>> last accessed 4 May 2021

Qazvinian V and others, 'Rumor has it: Identifying Misinformation in Microblogs' (2011) *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* 1589

Queller S, Schell T, and Mason W, 'A Novel View of Between-Categories Contrast and Within-Category Assimilation' (2006) 91 (3) *J Personality and Social Psychology* 406

- Rahman K S, 'The New Utilities: Private Power, Social Infrastructure, and the Revival of the Public Utility Concept' (2018) 39 Cardozo Law Review 1621
- Ransom M R, Kast C, and Shelly R K, 'Self-Enhancement, Self-Protection and Ingroup Bias' (2015) 23 Current Research in Social Psychology 56
- Ratkiewicz J and others, 'Detecting and Tracking Political Abuse in Social Media' (2011) Proceedings of the Fifth International Association for the Advancement of Artificial Intelligence Conference on Weblogs and Social Media 297
- Ray D G and Matschke C, 'Cross-group recognition bias generalizes to diverse non-face representations of digital identity' (2012) 48 J Experimental Social Psychology 387
- Read L E, *'I, Pencil: My Family Tree' as told to Leonard E. Read* (Foundation for Economic Education 1958)
- Reed C, 'How to make bad law: Lessons from Cyberspace' (2010) 73(6) Modern Law Review 903
- Reed C, *Making Laws for Cyberspace* (Oxford UP 2012)
- Reid A and Dotto C, 'Thousands of misleading Conservative ads side-step scrutiny thanks to Facebook policy' (First Draft, 6 December 2019)
<<https://firstdraftnews.org/latest/thousands-of-misleading-conservative-ads-side-step-scrutiny-thanks-to-facebook-policy/>> accessed 28 July 2022
- Renton D, 'Corbyn's Suspension' (London Review of Books, 31 August 2021)
<<https://www.lrb.co.uk/blog/2021/august/corbyn-s-suspension>> accessed 01 November 2022
- Republican National Committee, 'Principles for American Renewal' (2020)
<<https://www.gop.com/principles-for-american-renewal>> last accessed 4 May 2021
- Rex B F, 'Liability of Telegraph Companies for Fraud, Accident, Delay and Mistakes in the Transmission and Delivery of Messages' (May 1884) 32(5) American Law Register 281
- Roberts S T, 'Commercial Content Moderation: Digital Laborers' Dirty Work' (2016) Media Studies Publications, Paper 12 <<https://ir.lib.uwo.ca/commpub/12>>

Robertson K, 'New York Post Published Hunter Biden Report Amid Newsroom Doubts' (*The New York Times*, 18 October 2020)

<<https://www.nytimes.com/2020/10/18/business/media/new-york-post-hunter-biden.html>> accessed 01 November 2022

Robinson P, 'The CNN effect: can the news media drive foreign policy?' (April 1999) 25(2) *Review of International Studies* 301

Rosenberg M, Confessore N, and Cadwalladr C, 'How Trump Consultants Exploited the Facebook Data of Millions' (*The New York Times*, 17 March 2018)

<<https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html>> accessed 28 August 2022

Ross B and others, 'Measuring the Reliability of Hate Speech Annotations: The Case of the European Refugee Crisis' (2016) 17 *Proceedings of the Third Workshop on Natural Language Processing for Computer-Mediated Communication* 6

Ross W G, *The Chief Justiceship of Charles Evans Hughes, 1930-1941* (U of S Carolina P 2007)

Roth D, 'The infinite scroll' (*Columbia Journalism Review*, 04 March 2020)

<https://www.cjr.org/first_person/the-infinite-scroll.php> accessed 28 July 2022

Rothschild Z K and others, 'Displacing blame over the ingroup's harming of a disadvantaged group can fuel moral outrage at a third-party scapegoat' (2013) 49 *J Experimental Social Psychology* 898

Royal Mail Group, 'Our story' (2022) <<https://www.royalmailgroup.com/en/about-us/our-story/>> accessed 6 March 2022

Ryan J, 'Labour poses an existential threat to British Jews' (*The Jewish Chronicle*, 20 February 2019) <<https://www.thejc.com/comment/comment/i-cannot-remain-in-a-party-that-poses-an-existential-threat-to-british-jews-1.480314>> accessed 29 October 2022

Salmon F, 'Trust in media hits new crisis low' (*Axios*, 21 January 2021)

<<https://www.axios.com/2021/01/21/media-trust-crisis>> accessed 28 July 2022

Samples J and Matzko P, 'Social Media Regulation in the Public Interest: Some Lessons from History' (4 May 2020) Knight First Amendment Institute, Columbia University

<https://knightcolumbia.org/content/social-media-regulation-in-the-public-interest-some-lessons-from-history>> accessed 6 March 2022

Schain C and others, 'Looking at the actor's face: Identity cues and attentional focus in false memories of action performance from observation' (2012) 48 J Experimental Social Psychology 1201

Scherer A M, Windschitl P D, and Smith A R, 'Hope to be right: Biased information seeking following arbitrary and informed predictions' (2013) 49 J Experimental Social Psychology 106

Schwartz A B, 'The Infamous "War of the Worlds" Radio Broadcast Was a Magnificent Fluke' (Smithsonian Magazine, 06 May 2015)
<<https://www.smithsonianmag.com/history/infamous-war-worlds-radio-broadcast-was-magnificent-fluke-180955180/>> accessed 28 July 2022

Schleifer T, 'Inside the new \$65 million push from progressives to compete with conservative media' (Vox, 23 February 2021)
<<https://www.vox.com/platform/amp/recode/22296522/acronym-democrats-courier-progressive-news-tara-mcgowan-project-for-good-information>> accessed 28 July 2022

Seger C R, Smith E R, and Mackie D M, 'Subtle activation of a social categorization triggers group-level emotions' (2009) 45 J Experimental Psychology 460

Sforza L, 'Blue check chaos: Twitter policy spurs confusion and scramble to prove authenticity' (TheHill.com, 21 April 2023)
<<https://thehill.com/policy/technology/3963163-blue-check-chaos-twitter-policy-spurs-confusion-and-scramble-to-prove-authenticity/>> accessed 30 November 2023

Shearer E, 'Social media outpaces print newspapers in the U.S. as a news source' (Pew Research Center, 10 December 2018) <<https://www.pewresearch.org/fact-tank/2018/12/10/social-media-outpaces-print-newspapers-in-the-u-s-as-a-news-source/>> accessed 28 July 2022

Shakespeare W, *Henry VIII* (Arden, 3rd edn, 2000)

— — *The Merchant of Venice* (Arden, 2nd edn, 1964)

- Shank D B and Burns A, 'Comparing Groups' Affective Sentiments to Group Perceptions' (2018) 26 *Current Research in Social Psychology* 55
- Shannon C E, *A Mathematical Theory of Communication* (U Chicago P, 1949)
- Sharrett C, '9/11, the Useful Incident, and the Legacy of the Creel Committee' (2004) 43(4) *Cinema Journal* 125
- Shirky C, 'The Political Power of Social Media' (*Foreign Affairs*, January/February 2011) <<https://www.foreignaffairs.com/political-power-social-media>> accessed 28 July 2022
- Silberstein-Loeb J, 'The Structure of the News Market in Britain, 1870–1914' (Winter 2009) 83 *Business History Review* 759
- Silverman C, Mac R, and Dixit P, "'I Have Blood on My Hands": A Whistleblower Says Facebook Ignored Global Political Manipulation' (*BuzzFeed*, 14 September 2020) <<https://www.buzzfeednews.com/article/craigsilverman/facebook-ignore-political-manipulation-whistleblower-memo>> accessed 28 July 2022
- Sim J J and others, 'How social identity shapes the working self-concept' (2014) 55 *J Experimental Social Psychology* 271
- Sinha J and Naykankuppam D, 'The canny social judge: Predicting others' attitudes from sparse information' (2014) 53 *J Experimental Social Psychology* 145
- Sissons M and Warofka A, 'An Update on Facebook's Human Rights Work in Asia and Around the World' (Meta, 12 May 2020) <<https://about.fb.com/news/2020/05/human-rights-work-in-asia/>> accessed 28 July 2022
- Skibinski M, 'Special Report: The U.S. Best and Worst of 2021' (*NewsGuard*, 29 December 2021) <<https://www.newsguardtech.com/special-reports/special-report-the-u-s-best-and-worst-of-2021/>>
- Smalley S, 'What we know about the DHS' Disinformation Governance Board' (Poynter Institute, 05 May 2022) <<https://www.poynter.org/reporting-editing/2022/what-is-department-homeland-security-disinformation-governance-board/>> accessed 15 November 2022

- Smith B, Donfried K, and LeBlanc D, 'Multi-Stakeholder Insights: A Compendium on Countering Election Interference' (Alliance for Securing Democracy, 14 April 2021) <<https://www.canada.ca/content/dam/di-id/documents/compendium-eng.pdf>> accessed 28 July 2022
- Smith E R, Seger C R, and Mackie D M, 'Can Emotions Be Truly Group Level? Evidence Regarding Four Conceptual Criteria' (2007) 93(3) J Personality and Social Psychology 431
- Solon O and Laughland O, 'Cambridge Analytica closing after Facebook data harvesting scandal' (*The Guardian*, 02 May 2018) <<https://www.theguardian.com/uk-news/2018/may/02/cambridge-analytica-closing-down-after-facebook-row-reports-say>> accessed 28 August 2022
- Solove D, *The Future of Reputation: Gossip, Rumor, and Privacy on the Internet* (Yale UP 2007)
- Solsman J E and Wong Q, 'Facebook isn't taking political ads "for the money," COO Sandberg says' (*CNet.com*, 22 October 2019) <<https://www.cnet.com/news/facebook-isnt-taking-political-ads-for-the-money-coo-sheryl-sandberg-says/>> accessed 28 July 2022
- Sonne P, Kranish M, and Viser M, 'The gas tycoon and the vice president's son: The story of Hunter Biden's foray into Ukraine' (*Washington Post*, 28 September 2019) <https://www.washingtonpost.com/world/national-security/the-gas-tycoon-and-the-vice-presidents-son-the-story-of-hunter-bidens-foray-in-ukraine/2019/09/28/1aadff70-dfd9-11e9-8fd3-d943b4ed57e0_story.html> accessed 01 November 2022
- Sony, 'Product & Technology Milestones: Personal Audio' (2022) <<https://www.sony.com/en/SonyInfo/CorporateInfo/History/sonyhistory-e.html>> accessed 05 July 2022
- Springhall J, '"A Life Story for the People"? Edwin J. Brett and the London "Low-Life" Penny Dreadfuls of the 1860s' (Winter 1990) *Victorian Studies* 223
- Stahl L, 'How does YouTube handle the site's misinformation, conspiracy theories, and hate?' (*60 Minutes*, 01 December 2019) <<https://www.cbsnews.com/news/is-youtube-doing-enough-to-fight-hate-speech-and-conspiracy-theories-60-minutes-2019-12-01/>> accessed 28 July 2022

Stocking G and others, 'The Role of Alternative Social Media in the News and Information Environment' (Pew Research Center, 06 October 2022)

<<https://www.pewresearch.org/journalism/2022/10/06/the-role-of-alternative-social-media-in-the-news-and-information-environment/>> accessed 25 April 2023

Stone J, 'Anti-Corbyn Labour officials worked to lose general election to oust leader, leaked dossier finds' (*Independent*, 13 April 2020)

<<https://www.independent.co.uk/news/uk/politics/labour-leak-report-corbyn-election-whatsapp-antisemitism-tories-yougov-poll-a9462456.html>> accessed 28 August 2022

Strupp-Levitsky M and others, 'Moral "foundations" as the product of motivated social cognition: Empathy and other psychological underpinnings of ideological divergence in "individualizing" and "binding" concerns' (2020) 15 (11) PLoS ONE: e0241144

<<https://doi.org/10.1371/journal.pone.0241144>> accessed 15 October 2023

Sudhaman A, 'Davos: Edelman Wants PR to Shift From Advocacy to Education in Fake News Era' (Provoke Media, 23 Jan 2018)

<<https://www.provokemedia.com/latest/article/davos-edelman-wants-pr-to-shift-from-advocacy-to-education-in-fake-news-era>> accessed 28 July 2022

Taibbi M, 'Who Will Fix Facebook?' (Rolling Stone, 26 November 2018)

<<https://www.rollingstone.com/politics/politics-features/who-will-fix-facebook-759916/>> accessed 28 July 2022

Taibbi M, 'A Biden Appointee's Troubling Views On The First Amendment' (*TK News*, 23 March 2021)

<<https://taibbi.substack.com/p/a-biden-appointees-troubling-views>> accessed 28 July 2022

Taibbi M, 'Twitter Files' 18 (16 December 2022)

<<https://twitter.com/mtaibbi/status/1603857603867598859?s=20&t=2OrkAnvOz9O0QSMHh-ow4A>> accessed 04 January 2023

Taibbi M, 'Twitter Files' 24 (16 December 2022)

<<https://twitter.com/mtaibbi/status/1603857632091148288?s=20&t=2OrkAnvOz9O0QSMHh-ow4A>> accessed 04 January 2023

Tajfel H, 'Social Psychology of Intergroup Relations' (1982) 33 Annual Review of Psychology

- Tajfel H and others, 'Social Categorization and Intergroup Behaviour' (1971) 1(2) Eur J Social Psychology 149
- Tan W and others, 'Social-Network-Sourced Big Data Analytics' (September/October 2013) IEEE Internet Computing 62
- Tarrant M and others, 'Social identity and perceptions of torture: It's moral when we do it' (2012) 48 J Experimental Social Psychology 513
- Taub A and Fisher M, 'Where Countries Are Tinderboxes and Facebook Is a Match' (*New York Times*, 21 April 2018) <<https://www.nytimes.com/2018/04/21/world/asia/facebook-sri-lanka-riots.html>> accessed 28 August 2022
- Taylor D M, *The Quest for Identity: From Minority Groups to Generation Xers* (Praeger 2002)
- Taylor D M and Moghaddam F M, *Theories of Intergroup Relations: International Social Psychological Perspectives* (2nd edn, Praeger 1994)
- Taylor W S, 'Is Truth Individual or Social?' (August 1935) J Social Psychology 348
- Timberg C and others, 'Secret campaign to use Russian-inspired tactics in 2017 Ala. Election stirs anxiety for Democrats' (*Washington Post*, 06 January 2019) <https://www.washingtonpost.com/business/technology/secret-campaign-to-use-russian-inspired-tactics-in-2017-alabama-election-stirs-anxiety-for-democrats/2019/01/06/58803f26-0400-11e9-8186-4ec26a485713_story.html> accessed 28 July 2022
- Timberg C, Romm T, and Davis A C, 'Researcher whose firm wrote report on Russian interference used questionable online tactics during Ala. Senate race' (*Washington Post*, 19 December 2018) <<https://www.washingtonpost.com/technology/2018/12/19/researcher-affiliated-with-russian-interference-senate-report-used-questionable-online-tactics-during-alabama-senate-race/>> accessed 28 July 2022
- Timberg C, Viser M, & Hamburger T, 'Here's how The Post analyzed Hunter Biden's laptop' (*Washington Post*, 30 March 2022) <<https://www.washingtonpost.com/technology/2022/03/30/hunter-biden-laptop-data-examined/>> accessed 01 November 2022

- Tormala Z L, Clarkson J J, and Petty R E, 'Resisting Persuasion by the Skin of One's Teeth: The Hidden Success of Resisted Persuasive Messages' (2006) 91 J Personality and Social Psychology 423
- Travaglino G A and others, 'We are all in this together: The role of individuals' social identities in problematic engagement with video games and the internet' (2020) 59 (2) British J Social Psychology 522
- Tufekci Z and Wilson C, 'Social Media and the Decision to Participate in Political Protest: Observations from Tahrir Square' (2012) 62 J Communications 363
- TV Parental Guidelines Monitoring Board, 'Our History' (2021)
<<http://www.tvguidelines.org/aboutUs.html>> accessed 05 July 2022
- Twight C, 'What Congressmen Knew and When They Knew It: Further Evidence on the Origins of U.S. Broadcasting Regulation' (June 1998) 95(3) Public Choice 247
- Twitter, 'World Leaders on Twitter' (05 January 2018)
<https://blog.twitter.com/en_us/topics/company/2018/world-leaders-and-twitter>
accessed 26 August 2022
- Twitter, 'Civic integrity policy' (October 2020) <on file>
- Twitter, 'Our approach to blocking links' (July 2020) <<https://help.twitter.com/en/safety-and-security/phishing-spam-and-malware-links>> accessed 28 July 2022
- Twitter, 'Distribution of hacked materials policy' (October 2020)
<<https://help.twitter.com/en/rules-and-policies/hacked-materials>> accessed 01 November 2022
- Twitter, 'Permanent suspension of @realDonaldTrump' (8 January 2021)
<https://blog.twitter.com/en_us/topics/company/2020/suspension> accessed 28 July 2022
- Twitter, 'Civic integrity policy' (October 2021) <<https://help.twitter.com/en/rules-and-policies/election-integrity-policy>> accessed 28 July 2022
- Twitter, 'About: Our Purpose' (2022) <<https://about.twitter.com/en/who-we-are/our-company>> accessed 28 July 2022

Twitter, Inc., *Fiscal Year 2018 Annual Report* (2019)

<<https://investor.twitterinc.com/financial-information/annual-reports/default.aspx>>

Twitter Safety, 'Disclosing networks of state-linked information operations' (23 February 2021) <https://blog.twitter.com/en_us/topics/company/2021/disclosing-networks-of-state-linked-information-operations-> accessed 28 July 2022

Twitter Safety, 'Expanding our policies to further protect the civic conversation' (10 September 2020) <https://blog.twitter.com/en_us/topics/company/2020/civic-integrity-policy-update> accessed 26 August 2022

Twitter Safety (01 October 2020)

<<https://twitter.com/TwitterSafety/status/1311462538056544258?s=20>> (accessed 28 August 2020)

Twitter Support (14 March 2021)

<<https://twitter.com/twittersupport/status/1371235587718582274>> accessed 01 December 2022

UK Parliament, 'General Elections' (2021)

<<https://www.parliament.uk/about/how/elections-and-voting/general/>> accessed 28 April 2021

UN General Assembly 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression' (6 September 2016) UN Doc A/71/373

UN General Assembly 'Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression' (6 April 2018) UN Doc A/HRC/38/35

UN Human Rights Committee, 'General comment No. 37 (2020) on the right of peaceful assembly (article 21)' (17 September 2020) CCPR/C/GC/37

UN Special Rapporteur on Freedom of Opinion and Expression, OSCE Representative on Freedom of the Media, OAS Special Rapporteur on Freedom of Expression, and ACHPR Special Rapporteur on Freedom of Expression and Access to Information, 'Joint Declaration on Freedom of Expression and "Fake News", Disinformation and Propaganda' FOM.GAL/3/17 (Vienna, 3 March 2017)

USA TODAY, 'USA TODAY removed 23 stories from website, other platforms following audit of reporter's work' (*USA TODAY*, 16 June 2022)

<<https://eu.usatoday.com/story/news/2022/06/16/usa-today-audit-reporter/7647731001/>> accessed 28 July 2022

US Census Bureau, 'Annual Estimates of the Resident Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2019' (April 2021)

<<https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html>> last accessed 05 October 2021

US Census Bureau, Current Population Survey: Annual Social and Economic Supplement 2019 (April 2020) <<https://www.census.gov/data/tables/2019/demo/age-and-sex-composition.html>> last accessed 05 October 2021

US Census Bureau, 'QuickFacts: United States' (July 2019)

<<https://www.census.gov/quickfacts/fact/table/US/PST045219>> last accessed 05 October 2021

US Department of Homeland Security, 'Following HSAC Recommendations, DHS terminates Disinformation Governance Board' (24 August 2022)

<<https://www.dhs.gov/news/2022/08/24/following-hsac-recommendation-dhs-terminates-disinformation-governance-board>> accessed 15 November 2022

US Department of Justice, 'Section 230 – Nurturing Innovation or Fostering Unaccountability?' (June 2020) <<https://www.justice.gov/file/1286331/download>> accessed 28 July 2022

US Federal Trade Commission, 'FTC Imposes \$5 Billion Penalty and Sweeping New Privacy Restrictions on Facebook' (24 July 2019) <<https://www.ftc.gov/news-events/news/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions-facebook>> accessed 28 August 2022

US House Committee on Energy and Commerce, *Testimony of Jack Dorsey Chief Executive Officer Twitter, Inc.* (5 September 2018)

US House Committee on Energy and Commerce, 'E&C Committee Announces Hearing with Tech CEOs on the Misinformation and Disinformation Plaguing Online Platforms' (18 February 2021) <<https://energycommerce.house.gov/newsroom/press-releases/ec->

[committee-announces-hearing-with-tech-ceos-on-the-misinformation-and](#)> accessed 28 August 2022

US House Committee on the Judiciary and Permanent Select Committee on Intelligence, Joint Letter to Sec'y of State Antony Blinken (20 April 2023)

<https://judiciary.house.gov/sites/evo-subsites/republicans-judiciary.house.gov/files/evo-media-document/2023-04-20-jdj-mt-to-blinken-re-public-statement-on-hunter-biden-emails_0.pdf> accessed 01 December 2023

US Internal Revenue, *Statistics of Income* (GPO 1919) 24 <<https://www.irs.gov/pub/irs-soi/17soirepar.pdf>> accessed 6 March 2022

US National Archives and Records Administration, 'Alien and Sedition Acts (1798),' (2022) Milestone Documents <<https://www.archives.gov/milestone-documents/alien-and-sedition-acts>> accessed 6 March 2022

US Postal Service, *The United States Postal Service: An American History* (Washington, D.C. 2020)

US Senate Committee on the Judiciary and US Senate Committee on Commerce, Science and Transportation, *Testimony of Mark Zuckerberg Chairman and Chief Executive Officer, Facebook* (10 April 2018)

US Senate Select Committee on Intelligence, *Testimony of Jack Dorsey Chief Executive Officer Twitter, Inc.* (5 September 2018)

Vallone R P, Ross L, and Lepper M R, 'The Hostile Media Phenomenon: Biased Perception and Perceptions of Media Bias in Coverage of the Beirut Massacre' (1985) 49 (3) *J of Personality and Social Psychology* 57

Van Swol L M, 'Extreme members and group polarization' (2009) 4 (3) *Social Influence* 185

Viser M, Hamburger T, and Timberg C, 'Inside Hunter Biden's multimillion-dollar deals with a Chinese energy company' (*Washington Post*, 30 March 2022)
<<https://www.washingtonpost.com/politics/2022/03/30/hunter-biden-china-laptop/>> accessed 01 November 2022

Vitriol J A and others, 'Ideological uncertainty and investment of the self in politics' (2019) 82 *J Experimental Social Psychology* 85

Warner M, 'Warner: It's past time to learn from failures, adapt our laws to the internet age' (01 August 2018)

<<https://www.warner.senate.gov/public/index.cfm/2018/8/warner-it-s-past-time-to-learn-from-failures-adapt-our-laws-to-the-internet-age>> accessed 28 July 2022

Warner M, 'Potential Policy Proposals for Regulation of Social Media and Technology Firms' (White Paper, August 2018) <<https://www.scribd.com/document/385137394/MRW-Social-Media-Regulation-Proposals-Developed>> accessed 28 July 2022

Warren E, 'Fighting digital disinformation' (Warren for President, 2020) <<https://elizabethwarren.com/plans/fighting-digital-disinformation>> accessed 28 July 2022

Waseem Z, 'Are You a Racist or Am I Seeing Things? Annotator Influence on Hate Speech Detection on Twitter' (2016) Association for Computational Linguistics, Proceedings of 2016 Workshop on Natural Language Processing and Computational Social Science 138

Washington Post Editorial Board, 'The Hunter Biden story is an opportunity for a reckoning' (*Washington Post*, 03 April 2022) <<https://www.washingtonpost.com/opinions/2022/04/03/hunter-biden-story-is-an-opportunity-reckoning/>> accessed 01 November 2022

Waterson J, 'YouTube blocks history teachers uploading archive videos of Hitler' (*The Guardian*, 06 June 2019) <<https://www.theguardian.com/technology/2019/jun/06/youtube-blocks-history-teachers-uploading-archive-videos-of-hitler>> accessed 24 August 2022

Wauters E, Lievens E, and Valcke P, 'Social Networking Sites' Terms of Use: Addressing Imbalances in the User-Provider Relationship through Ex Ante and Ex Post Mechanisms' (2014) 5 Journal of Intellectual Property, Information Technology and Electronic Commerce Law 139

Wells C E, 'The Promise and Peril of Protest in the Internet Era' (2012) University of Missouri Legal Studies Research Paper Series 2012-23

Wilton Park, '#FakeNews: innocuous or intolerable?' (Wilton Park Report WP1542, 15-17 February 2017)

Winston B, *Media technology and society: a history* (Routledge 1998)

Winston B and Winston M, *The Roots of Fake News: Objecting to Objective Journalism* (Routledge 2021)

Woodrow Wilson House, 'Election: 1916' (19 November 2020)

<<https://www.woodrowwilsonhouse.org/election-1916/>> accessed 6 March 2022

Wu T, 'Is the First Amendment Obsolete?' in *The Free Speech Century* (eds Geoffrey R Stone & Lee C Bollinger, Oxford UP 2019) 272

Xiao Y J and Van Bavel J J, 'Sudden shifts in social identity swiftly shape implicit evaluation' (2019) 83 J Experimental Social Psychology 55

York J C and McSherry C, 'Content Moderation is Broken: Let Us Count the Ways' (Electronic Frontier Foundation, 29 April 2019)

<<https://www.eff.org/deeplinks/2019/04/content-moderation-broken-let-us-count-ways>> accessed 28 July 2022

Younis M, 'China, Russia Images in U.S. Hit Historic Lows' (Gallup, 01 March 2021)

<<https://news.gallup.com/poll/331082/china-russia-images-hit-historic-lows.aspx>> accessed 28 July 2022

YouTube Help, 'Impersonation Policy' (Google, 2022)

<https://support.google.com/youtube/answer/2801947?hl=en&ref_topic=9282365> (accessed 28 August 2022)

YouTube Help, 'Misinformation policies' (Google, 2022)

<<https://support.google.com/youtube/answer/10834785>> accessed 28 August 2022

YouTube Help, 'Elections misinformation policies' (Google, 2022)

<<https://support.google.com/youtube/answer/10835034>> accessed 28 August 2022

The YouTube Team, 'Continuing our work to improve recommendations on YouTube' (25 January 2019)

<<https://blog.youtube/news-and-events/continuing-our-work-to-improve/>> accessed 24 August 2022

Zhang C and Conrad F, 'Speeding in Web Surveys: The tendency to answer very fast and its association with straightlining' (2014) 8(2) Survey Research Methods 127

Zmigrod L, 'A Psychology of Ideology: Unpacking the Psychological Structure of Ideological Thinking' (2022) 17 (4) Perspectives on Psychological Science 1072

Zmigrod L and Goldenberg A, 'Cognition and Emotion in Extreme Political Action: Individual Differences and Dynamic Interactions' (2021) 30 (3) Current Directions in Psychological Science 218

Zoellner D, 'Twitter users flag bug where accounts are banned if they tweet "Memphis"' (*Independent*, 14 March 2021) <<https://www.independent.co.uk/tech/twitter-accounts-ban-memphis-b1817064.html>> accessed 28 July 2022