Misinformation and Disinformation: the Issue, the Marketplace of Ideas, and the British and American Approaches

KIERAN JON SEWELL

Misinformation and Disinformation: the Issue, the Marketplace of Ideas, and the British and American Approaches

Kieran Jon Sewell, MJur

Abstract

This thesis concerns the shortcomings of free speech theory in addressing misinformation and disinformation. The thesis principally aims to demonstrate the inapplicability of justifications for freedom of expression grounded in the search for the truth to the problem of misinformation and disinformation. It does so by examining whether individuals are able to perform 'truth-seeking' in relation to misinformation and disinformation given certain psychological biases and technological developments of which misinformation and disinformation take advantage. The marketplace of ideas is one justification for freedom of expression based upon the discovery of the truth, which considers 'the best test of truth [to be] the power of the thought to get itself accepted in the competition of the market.' This thesis argues that preventing misinformation and disinformation from entering this market should generally be the preferred solution to the issue, though others may be appropriate where certain speech interests apply. This is demonstrated through analysis of legislation and proposals that address misinformation and disinformation directly, as well as the technologies used to disseminate such speech.

I certify that this thesis has been composed by me, that the work contained in it is my own and that it has not been submitted for any other degree or professional qualification.

Kieran Jon Sewell, 23 June 2023.

Table of Contents

Acknowledgements

I am extremely grateful to Dr Dimitrios Kagiaros. Without your unending and unequivocal support, I am not sure I would have embarked on this path to begin with. I am sure however that without your feedback I would not have overcome many of the barriers I encountered, let alone have spotted them. Thank you.

I would also like to thank Katie Morris, Eshbal Geifman and Elizabeth Tracey for bearing with me as I endlessly ranted and complained about the law, my inability to understand the law, and whatever else happened to enter my mind when I was trying to talk to them about the law.

Table of Cases

United Kingdom

- *Lee v Ashers Baking Company Ltd and Others (Northern Ireland)* [2018] UKSC 49.
- *Percy v Director of Public Prosecutions* [2001] EWHC 1125 (Admin).
- *RT (Zimbabwe)* [2012] UKSC 38.
- *Spiller v Joseph* [2010] UKSC 53.

United States of America

- *Abrams v US* [1919] 250 US 616.
- *Brandenburg v Ohio* [1969] 395 US 444.
- *Buckley v Valeo* [1976] 424 US 1.
- *Citizens United v Federal Election Commission* [2010] 558 US 310.
- *Cohen v California* [1971] 403 US 15.
- *Doe v Gonzales* [2005] 386 FSupp 2d 66 (D. Conn.).
- *Doe v Reed* [2010] 561 US 186.
- *Hurley v Irish-American Gay, Lesbian and Bisexual Group of Boston, Inc.* [1995] 515 US 557.
- *Konigsberg v State Bar of California* [1961] 366 US 36.
- *Linmark Associates, Inc. v Township of Willingboro* [1977] 431 US 85.
- *Lorillard Tobacco Co v Reilly* [2001] 533 US 525.
- *McIntyre v Ohio Elections Commission* [1995] 514 US 334.
- *National Electrical Manufacturers Association v Sorrell* [2001] 272 F3d 104 (2d Cir).
- *New York State Restaurant Association v New York City Board of Heath* [2009] 556 F3d 114, 132 (2d Cir).
- *New York Times v Sullivan* [1964] 376 US 254.
- *Pacific Gas & Electric Co. v Public Utilities Commission* [1986] 475 US 1.
- *Police Department of the City of Chicago v Mosley* [1972] 408 US 92.

- *Reed v Town of Gilbert, Arizona* [2015] 576 US 155.
- *Rosenberger v Rector and Visitors of University of Virginia* [1995] 515 US 819.
- *Roth v US* [1957] 354 US 476.
- *Sable Communications v Federal Communications Commission* [1989] 492 US 115.
- *Talley v State of California* [1960] 362 US 60.
- *Texas v Johnson* [1989] 491 US 397.
- *Turner Broadcasting System v Federal Communications Commission* [1994] 512 US 622.
- *US v Alvarez* [2012] 567 US 1.
- *Ward v Rock Against Racism* [1989] 491 US 781.
- *Watchtower Bible and Tract Society of New York v Village of Stratton* [2002] 536 US 150.
- *Whitney v California* [1927] 274 US 357.
- *Williams-Yulee v Florida Bar* [2015] 575 US 433.

<u>European Court of Human Rights</u>

- *Ahmet Yıldırım v Turkey* App. No.3111/10.
- *Ali Gürbüz v Turkey* App. No.52497/08.
- *Altuğ Taner Akçam v Turkey* App. No.27520/07.
- *Andrushko v Russia* 4260/04.
- *Animal Defenders International v UK*, App. No.48876/08.
- *Autronic AG v Switzerland* App. No.12726/87.
- *Barthold v Germany* App. No.8734/79.
- *Bayev and Others v Russia* App. Nos.67667/09, 44092/12, 56717/12.
- *Bladet Tromsø and Stensaas v. Norway* App. No.21980/93.
- *Bowman v UK* App. No.24839/94.
- *Breyer v Germany* App. No.50001/12.
- *Buscarini and Others v San Marino* App. No.24645/94.
- *Casado Coca v Spain* App. No.15450/89.

- *Castells v Spain* App. No.11798/85.
- *Cengiz and Others v Turkey* App. Nos.48226 and 14027/11.
- *Delfi AS v Estonia* App. No.64569/09.
- *Einarsson v Iceland* App. No.24703/15.
- *ES v Austria*, App. No.38450/12.
- *Gillberg v Sweden* App. No.41723/06.
- *Glor v Switzerland* App.No.13444/04.
- *Glor v Switzerland* App.No.13444/04.
- *Handyside v UK* App. No.5493/72.
- *Hirst v UK* App. No.74025/01.
- *Kandzhov v Bulgaria* App. No.68924/01.
- *Karatas v Turkey* App. No.23168/94.
- *Kokkinakis v Greece* App. No.14307/88.
- *KU v Finland* App. No.2872/02.
- *Larissis and Others v Greece* App. No.23372/94.
- *Lewandowska-Malec v Poland* App. No.39660/07.
- *Lingens v Austria* App. No.9815/82.
- *Lingens v Austria* App. No.9815/82.
- *Magyar Helsinki Bizottság v Hungary* App. No.18030/11.
- *Magyar Kétfarkú Kutya Párt v Hungary* App. No.201/17.
- *Marin Kostov v Bulgaria* App. No.13801/07.
- *Mariya Alekhina and Others v Russia* App. No.38004/12.
- *markt intern Verlag GmbH and Klaus Beermann* App. No.10572/83.
- *Matalas v Greece* App. No.1864/18.
- *Monnat v Switzerland* App. No.73604/01.
- *Mouvement Raëlien Suisse v Switzerland* App. No.16354/06.
- *Müller v Switzerland* App. No.10737/84.

- *Otegi Mondragon v Spain* App. No.2034/07.
- *Otto-Preminger-Institut v Austria* App. No.13470/87.
- *Perinçek v Switzerland* App. No.27510/08.
- *Polat v Turkey* App. No.23500/94.
- *Raichinov v Bulgaria* App. No.47579/99.
- *Sekmadienis Ltd v Lithuania* App. No.69317/1.
- *Sofranschi v Moldova* App. No.34690/05.
- *Standard Verlagsgesellschaft MBH v Austria (No 3)* App. No.39378/15.
- *The Sunday Times v UK (No.1)*, App. No.6538/74.
- *The Sunday Times v UK (No.1)*, App. No.6538/74.
- *Times Newspapers Ltd v UK (Nos 1 and 2)* App. Nos.3002/03 and 23676/03.
- *Toranzo Gomez v Spain* App. No.26922/14.
- *TV Vest AS and Rogaland Pensjonistparti v Norway* App. No.21132/05.
- *Vajnai v Hungary* App. No.33629/06.
- *Wanner v Germany* App. No.26892/12.
- *Wingrove v UK*, App. No.17419/90.
- *Wingrove v UK*, App. No.17419/90.
- *Young, James and Webster v UK* App. No.7601/76.
- *Zakharov v Russia* App. No.14881/03.
- *Ždanoka v Latvia* App. No.58278/00.

Chapter 1: Introduction

This thesis argues that strategies for dealing with misinformation and disinformation that are reliant upon discussion, "more speech" or "counterspeech" as remedies for falsity will be ineffective. "More speech" is not a solution to the harms that misinformation and disinformation present because discussion is not guaranteed to produce truth instead of falsity.

As such, this thesis argues that the preferred approach to the misinformation and disinformation issue is prevention of such speech in the first place. This solution is preferred because of the psychological and technological advantages misinformation and disinformation hold over other speech.

Upon the basis of these findings, this thesis examines the appropriateness of certain measures for addressing the misinformation and disinformation issue, and often argues for a preventative approach to be employed despite the free speech interests that are identified.

In lieu of "fake news," misinformation and disinformation are the commonly adopted terms to describe the speech with which this thesis is concerned. Misinformation can be defined as the inadvertent spreading of false factual information,[1] whereas disinformation is the deliberate spreading of false factual information.[2] These are the definitions that will be used for this thesis, however, other variations of the terms exist.[3] Given that the discussion in this thesis is generally applicable to both misinformation and disinformation, "anti-information" will be used to refer to both.

Chapter 1.1: Structure and Conclusions

The first aim of this thesis will be to demonstrate that a "more speech" approach to the anti-information issue cannot be relied upon. Preliminarily, it will need to be shown why the "truth-seeking" free speech

[1] Department for Digital, Culture, Media and Sport, *Online Harms* (CP 57, 2019) 23.
[2] ibid.
[3] See for example, Council of Europe, *Information Disorder: Towards an interdisciplinary framework for research and policy making* (September 2017) 20, available at <https://rm.coe.int/information-disorder-report-version-august-2018/16808c9c77> accessed 26 September 2022.

theories that advocate for "more speech" are the most applicable to the issue at hand, and this will be addressed in the methodology.

Chapter 2 will then question their applicability to factual debates, such as those that are distorted by anti-information. The ineffectiveness of a "more speech" approach to anti-information will be demonstrated by reference to the psychological advantages anti-information holds over other speech, and the role technology plays in the dissemination of anti-information.

Chapter 2 will conclude that "more speech" as a solution to falsity is ineffective, particularly because it is reliant upon having sufficient time. Chapter 2 will distinguish the real person that encounters anti-information from the rational person assumed by truth-seeking theories. The inapplicability of a "more speech" approach, the strengths anti-information holds over other speech, including those gained through new technologies, and the separation of the real person from the rational will lead to the conclusion that traditional remedies for falsity grounded in "truth-seeking" free speech theories cannot be relied upon to address the anti-information issue. Chapter 2 concludes that since it is questionable whether the "truth-seeking" theories are applicable to factual debates, and since we are not rational actors as those theories require, and even if we were, there may not be sufficient time to remedy falsity because new technologies mean we can be overwhelmed with anti-information, "more speech" is an absurd solution to the anti-information issue.

In discussing the role of technology in disseminating anti-information, this Chapter will highlight in particular the role of bots and targeted advertisements. In doing so, it will justify examining proposals that aim to address the role bots and targeted advertisements play in the dissemination of anti-information. Chapters 5 and 6 will therefore consider solutions to the anti-information issue in relation to these technologies.

The second aim of the thesis is to evaluate the effectiveness of statutes and proposals that address the anti-information issue in the US and the UK. Evaluating the likely effectiveness of proposals in the UK in light of the conclusions in Chapter 2 will help to determine the strengths and weaknesses of the UK's anti-information strategy. Evaluating statutes and proposals in the US can provide further options for

the UK, and can highlight issues that may arise either in ensuring the solution is effective or with regards to certain speech interests.

Before considering the effectiveness of different proposals in Chapters 4, 5 and 6, Chapter 3 will briefly discuss types of speech-related solutions (that may be) used to counter anti-information. Chapter 3, and the rest of the thesis, considers only speech-related solutions. Although other measures, particularly educative approaches such as media literacy measures,[4] could certainly be useful this thesis is limited to speech-related measures by virtue of the analysis in Chapter 2 and its subsequent greater utility with regards to speech-related measures.

Ultimately, measures that do not impact speech, like educative measures, are likely not an appropriate solution to the anti-information issue because if their effectiveness is dependent on time, like "more speech," it would be an ineffective solution where people can be overwhelmed with anti-information by new technologies. This point will be clarified in Chapter 2.

Chapter 3 will conclude that a preventative approach – where restrictions on speech are in place to prevent individuals encountering anti-information – is preferred because the strength of psychological effects and the role of technology in disseminating anti-information, as discussed in Chapter 2, lend too great an advantage to anti-information over other speech. That is not to say that solutions such as "more speech," information correction and removal of, or limitation of access to anti-information have no value, just that prevention is preferrable.

Chapters 4, 5 and 6 evaluate different legislative solutions to the anti-information issue. Whilst, as mentioned, Chapters 5 and 6 concern bots and targeted advertisements respectively, Chapter 4 discusses legislative proposals touted as solutions to the anti-information issue.

The appropriateness of each measure for addressing the anti-information issue is considered in light of different speech interests after they are examined closely to consider their effectiveness. The conclusions reached as to their effectiveness as a measure to combat anti-information and

[4] See for example, Xizhu Xiao, Porismita Borah and Yan Su, 'The dangers of blind trust: Examining the interplay among social media news use, misinformation identification, and news trust on conspiracy beliefs' (2021) 30(8) Public Understanding of Science 977-992, 986.

appropriateness of countering anti-information whilst respecting freedom of expression is informed by the discussion in Chapter 2 and the view it reaches on "truth-seeking" justifications for freedom of expression. American measures are considered first, as they may inform the discussion of options available to the UK.

Chapter 4 discusses the Honest Ads Bill (US) and the Online Safety Bill (UK). Chapter 4 finds that the approaches employed by these instruments are generally "more speech" approaches, however, some of their provisions employ preventative approaches that could effectively address the anti-information issue. Their effectiveness is dependent upon enforcement, and whilst they address many of the psychological barriers and technological strengths of anti-information, they do not provide alternative ideas to the idea-consumer to replace their beliefs in anti-information.

As there is a lack of proposals with regards to bots and targeted advertisements in the UK, the compatibility solutions like those proposed in the US with the European Convention on Human Rights, particularly Article 10, freedom of expression, will be considered briefly in both Chapter 5 and 6.

Chapter 5 concerns bot identity disclosure – the requirement that a bot discloses that it is a bot – a measure enacted in California. It concludes that knowing something is a bot has questionable benefits and whilst a measure like bot identity disclosure could be an option in the UK, a stricter measure would be preferred. Though a general measure, such as a ban, would too heavily interfere with the relevant speech interests.

This is due to the ability of bots to overwhelm the idea-consumer, by flooding their social media feed with anti-information. Similarly, targeted advertisements can overwhelm the idea-consumer with advertisements based upon false information.

Therefore, Chapter 6 discusses the Banning Microtargeted Political Advertisements Bill and finds that such a measure could be employed in the UK. This Chapter finds that a similar measure, as a preventative measure, would be appropriate to be used in the UK, particularly because measures that prevent anti-information entering the marketplace of ideas in the first place are the preferred solution.

Chapter 1.2: Methodology, Comparators and Theoretical Basis

This part will explain the multiple methodologies employed throughout this thesis. Also, it will explain the choice of comparators in more detail. Additionally, it will justify the use of "truth-seeking" free speech theories as the theoretical basis upon which this research was conducted and explain how the differences between the law on freedom of expression in the UK and the US does not prevent a useful comparison being made, and instead facilitates the search for inspiration for legislative solutions of the anti-information issue.

Chapter 1.2.1: Socio-legal Research

Regarding the thesis' first aim – to demonstrate that a "more speech" approach to the anti-information issue cannot be relied upon – a socio-legal approach to research is undertaken. This approach is interdisciplinary – it seeks to inform legal ideas and practices by reference to the social sciences, uncovering law's social effects and its relationship with social structures.[5]

To understand the nature of the issue, it is necessary to approach it from an interdisciplinary angle. Understanding the impact anti-information has had on society requires sociological research. Addressing that impact is only achievable with an understanding of the psychological phenomena at play,[6] and is greatly aided by reference to studies in communication theory.[7]

When considering the anti-information issue from the perspective of free speech theory, research from different disciplines can inform the rationality of the assumptions made and arguments posited by

[5] DR Harris, 'The development of socio-legal studies in the United Kingdom' (1983) 3(3) Legal Studies 315-33, 315.
[6] For example, "confirmation bias." See Brendan Nyhan and Jason Reifler, 'When Corrections Fail: The Persistence of Political Misperceptions' (2010) 32(2) Political Behaviour 303-330.
[7] For example, PR Chamberlain, 'Twitter as a Vector for Disinformation' (2010) 9(1) Journal of Information Warfare 11-17, which explains a great deal about Twitter's infrastructure.

different perspectives on freedom of expression. This analysis establishes that the "truth-seeking" theories presently do not account for the reality of the anti-information issue.

Chapter 1.2.2: Comparative Methodology

For the second aim of this thesis – to examine statutes and proposals in the US and UK that tackle the anti-information issue in order to inform the UK's anti-information strategy – a comparison will be undertaken which employs a functionalist approach. A functionalist approach differs from other comparative approaches such as classificatory or historical. Whilst these concern the grouping of legal systems into families or the understanding of development of law over time,[8] a functionalist approach examines similar constitutional institutions that perform the same function or how the same function is performed by different legal rules.[9]

The functional method allows the comparatist to find which is the "better" solution to the particular problem they both address.[10] The search for different solutions to the same problem is one of finding 'the better law'[11] however, for a comparatist to take such an approach, they run the risk of asserting what approach is "best" without any actual data to support such a conclusion.[12] The assumption of 'similarity of economic and social reality' tends to be found in such searches.[13] Therefore, to avoid making such a mistake, this thesis will avoid asserting that any one solution to the misinformation and disinformation problem is the "best" and instead focus on highlighting the factors that make particular solutions more or less appropriate and effective in the subject States.

The function to be compared is the ability of legislation to counter the effects of anti-information, with respect to the free speech theories that will be discussed, and conclusions drawn on their applicability to the anti-information issue. This comparison is valuable because it reveals the desirability of the use

[8] Vicki Jackson, 'Comparative Constitutional Law: Methodologies' 2-4, in Michael Rosenfeld and András Sajó (eds), *The Oxford Handbook of Comparative Constitutional Law* (Oxford University Press 2012).
[9] ibid 7.
[10] Geoffrey Samuel, *An Introduction to Comparative Law Theory and Method* (Hart Publishing 2014) 67, cf James Gordley, 'The Functional Method' 107-119 in Pier Giuseppe Monateri (ed), *Methods of Comparative Law* (Edward Elgar 2012).
[11] Jacques Herbots, 'Interpretation of contracts' in Jan Smits (ed), *Elgar Encyclopedia of Comparative* Law (Edward Elgar 2006), pages 325-346.
[12] Samuel (n 10) 55.
[13] Samuel (n 10) 56.

of the different legislative acts and proposals – a conclusion which is validated by how the acts and proposals correspond with the interpretation of the "truth-seeking" theories uncovered under the thesis' first aim.

Chapter 1.2.3: Selection of Comparators

The first consideration in selecting subject states for the comparison is to determine whether few or many states should form the subjects of comparison. Practical considerations are relevant when determining methodological questions, hence, given the length of this thesis, fewer states were preferred to many. A large portion of this thesis is already dedicated to its first aim, therefore, to ensure the action taken by subject states is examined in significant detail, only two states have been selected. Nonetheless, there is value in selecting fewer states.[14]

Secondly, it must be determined how subject states are to be selected. Whilst there is arguably no go-to rule as to how subject states should be selected,[15] particularly because it is dependent on the research topic,[16] the objective of the comparison is determinative.[17] Where the objective is seeking legislative inspiration, states that have similar political and socio-economic backgrounds that have encountered similar problems would serve as good comparators.[18]

The United Kingdom, as the system seeking inspiration, is selected inherently and the United States has been selected for reasons given below.

United States

From a practical perspective, there are a few reasons why the US is an appropriate comparator. Principally, relevant sources that originate from both States are highly accessible,[19] and since they are

[14] Taavi Annus, 'Comparative Constitutional Reasoning: The Law and Strategy of Selecting the Right Arguments' (2004) 14(2) Duke Journal of Comparative & International Law 301-50, 341.
[15] Leontin-Jean Constantinesco, *Traité de Droit Comparé, Tome II: La Méthode Comparative* (Libraire Générale de Droit et de Jurisprudence 1974) 41-2.
[16] Konrad Zweigert and Hein Kötz, *An Introduction to Comparative Law* (Tony Weir tr, 3rd edn, Oxford University Press 1998) 41.
[17] Constantinesco (n 15) 110.
[18] Ted de Boer, 'Vergelijkenderwijs: de inspiratie van buitenlands recht' (1992) 123(6033) Weekblad Voor Privaatrecht Notariaat en Registratie 39-48, 47-48, cf Constantinesco (n 15) 42.
[19] Recent research has focused on both States following the 2016 US Presidential Election and the Brexit Referendum – see, Edda Humprecht, Frank Esser, and Peter Van Aelst, 'Resilience to Online Disinformation: A

written in English, understood by the author. Additionally, in seeking inspiration for the UK, the many state legislators within the US, alongside the federal legislator, offer many sources of unique approaches to the issue.

There are many similarities between the UK and US that make for a useful comparison. Their shared common law traditions and the fact that the American free speech tradition owes its intellectual origins to the UK,[20] makes for a good starting point. The similar social experience of anti-information in the UK and US that makes them good comparators.

As States where distrust in news media and the government is relatively high, anti-information leads to political polarisation.[21] Further, in both States, anti-information originating from politicians, often used to attack political enemies, is frequent.[22] Individuals in both States are particularly aware of the anti-information issue and have opinions on it because of 'news coverage of online misinformation and [because of] prominent politicians using the term to attack journalism.'[23] The important role that social media plays in facilitating participation in the political process in both States has been demonstrated,[24] whilst the use of social media disinformation campaigns by political parties in both States has also been observed.[25] Consequently, anti-information of this type is of particular concern and importance for research in both States.

Framework for Cross-National Comparative Research' (2020) 25(3) The International Journal of Press/Politics 493-516, 497.

[20] Stephen J Shapiro, 'Comparing Free Speech: United States v. United Kingdom' (1989) 19(2) University of Baltimore Law Forum 17-20, 19. See also Chapter 2.1.

[21] Edda Humprecht, 'Where "fake news" flourishes: a comparison across four Western democracies' (2019) 22(13) Information, Communication & Society 1973-88, 1984.

[22] ibid.

[23] Rasmus Kleis Nielsen and Lucas Graves, '"News you don't believe": Audience perspectives on fake news' (Reuters Institute for the Study of Journalism 2017) <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-10/Nielsen%26Graves_factsheet_1710v3_FINAL_download.pdf> accessed 14 April 2023.

[24] Magdalena Saldaña, Shannon C McGregor, and Homero Gil De Zúñiga, 'Social Media as a Public Space for Politics: Cross-National Comparison of News Consumption and Participatory Behaviours in the United States and the United Kingdom' (2015) 9 International Journal of Communication 3304-3326, 3316.

[25] Samantha Bradshaw and Philip N Howard, 'The Global Organization of Social Media Disinformation Campaigns' (2018) 71(1.5) Journal of International Affairs 23-32, 27.

However, the anti-information issue in the UK and US is not limited to politics. The most Twitter users discussing misinformation relating to COVID-19 were from the US, followed by the UK.[26] Additionally, social media discussion of the claim that vaccines cause autism has been found to be more prevalent for parents in the UK and US,[27] where social media is among the 'top resources used by parents in the vaccination decision-making process.'[28]

Since the 2016 US Presidential Election and the Brexit Referendum, anti-information in the UK and US has been linked to populism,[29] anti-intellectualism,[30] and the decline of democratic institutions.[31] The effects of anti-information have been heavily felt and well observed in each, as such, the US is an appropriate comparator to the UK for research on this issue.

Objects of Comparison

The pieces of (proposed) legislation chosen for comparison are insufficient for a comprehensive comparison. Quite simply, because of the US' size and the number of states within, there are too many relevant instruments to discuss them all in detail. Many proposals in different states are very similar, therefore, it is assumed that it would be more valuable to consider either singular pieces of legislation that will affect all the states the same (i.e., federal legislation) or to discuss legislation that offers a unique approach.

[26] Binxuan Huang and Kathleen M Carley, 'Disinformation and Misinformation on Twitter during the Novel Coronavirus Outbreak' <https://arxiv.org/pdf/2006.04278.pdf> accessed 14 April 2023.
[27] S Mo Jang, Brooke W Mckeever, Robert Mckeever, and Joon Kyoung Kim, 'From Social Media to Mainstream News: The Information Flow of the Vaccine-Autism Controversy in the US, Canada and the UK' (2019) 34(1) Health Communication 110-117, 115.
[28] ibid. See also, Susan Goldstein, Noni E MacDonald, Sherine Guirguis, 'Health communication and vaccine hesitancy' (2015) 33(34) Vaccine 4212-4214.
[29] Eiríkur Bergmann, 'Populism and the politics of misinformation' (2020) 21(3) The Journal of South African and American Studies 251-65.
[30] Matthew Motta, 'The Dynamics and Political Implications of Anti-Intellectualism in the United States' (2018) 46(3) American Politics Research 465-498.
Farhana Sultana, 'The false equivalence of academic freedom and free speech: Defending academic integrity in the age of white supremacy, colonial nostalgia, and anti-intellectualism' (2018) 17(2) ACME: An International Journal for Critical Geographies 228-57, 248.
Ajnesh Prasad, 'Denying Anthropogenic Climate Change: Or, How Our Rejection of Objective Reality Gave Intellectual Legitimacy to Fake News' (2019) 34(SI) Sociological Forum 1217-1234.
[31] W Lance Bennett and Steven Livingston, 'The disinformation order: Disruptive communication and the decline of democratic institutions' (2018) 33(2) European Journal of Communication 122-139.
Timo Harjuniemi, 'Post-truth, fake news and the liberal "regime of truth" – The double movement between Lippmann and Hayek' (2022) 37(3) European Journal of Communication 269-283.

Additionally, because addressing anti-information through law is an emerging area, the majority of the selected objects of comparison are proposals. This is an unavoidable practical limitation of the research topic. The length of this thesis and the decision to select objects of comparison for their uniqueness means that there can be no claim within that the conclusions drawn from the comparison are comprehensive. A more thorough comparison would require a larger project/thesis.

Chapter 1.2.4: Theoretical Basis

This section will explain the use of "truth-seeking" free speech theories as the theoretical basis for this thesis. Additionally, it will explain how differences between the British and American conceptions of freedom of expression, Article 10 of the European Convention on Human Rights (ECHR) and the First Amendment to the US Constitution, do not prevent a meaningful comparison, and instead facilitate the aims of the thesis.

"Truth-Seeking" Free Speech Theories

The right to freedom of expression is not limited to spoken and written word. It may include expression such as flag burning,[32] offensive speech,[33] and artwork.[34] "Expression" or "speech" could extend to nearly limitless acts, however, a principle of freedom of expression cannot be so broad without risking conflation with a general principle of liberty – without a reason for protecting freedom of expression, it 'is more a platitude than a principle.'[35] Therefore, a free speech theory is one which determines what speech is,[36] and what speech is to be afforded protection by demonstrating that that speech furthers a

[32] *Percy v Director of Public Prosecutions* [2001] EWHC 1125 (Admin), [27]. *Texas v Johnson* [1989] 491 US 397, 403 (United States of America).
[33] *Handyside v UK* App. No.5493/72, [49]. *Cohen v California* [1971] 403 US 15, 24-6 (United States of America). *Texas* (n 32) 414.
[34] *Müller v Switzerland* App. No.10737/84, [27]. *Hurley v Irish-American Gay, Lesbian and Bisexual Group of Boston, Inc.* [1995] 515 US 557, 569 (United States of America).
[35] Frederick Schauer, *Free Speech: A Philosophical Enquiry* (Cambridge University Press 1982) 6. See also, Stanley Fish, *There's No Such Thing As Free Speech: And It's a Good Thing, Too* (Oxford University Press 1994) 123.
[36] Eric Barendt, *Freedom of Speech* (Oxford University Press 2007) 7.

goal which exists outside of the right.[37] Oft-cited goals include self-fulfilment/self-realisation,[38] democratic participation/popular sovereignty,[39] and suspicion of government.[40]

However, this thesis will rely upon arguments that are justified by the discovery/identification of truth, simply because by taking issue with falsity, the search for truth is the most applicable goal of those available. Reliance upon the argument from democracy could exclude the role of anti-information relating to health from the discussion, reliance upon the argument from self-fulfilment may not pay sufficient attention to the harms anti-information has on democratic society, and reliance upon suspicion of government is better explained as an argument against government regulation rather than one which justifies freedom of expression,[41] and subsequently offers little guidance on whether action taken against anti-information is compliant with the right or not. Truth on the other hand does not exclude any form of expression and allows for analysis of all aspects of the issue. The "truth-seeking" theories are outlined in Chapter 2.

Article 10 and the First Amendment

At present in the UK, the Human Rights Act 1998 gives domestic effect to the ECHR. Article 10 of the Convention, 'Freedom of Expression' is comprised of two paragraphs. Paragraph 1 makes provision for the right which includes 'freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers.' Paragraph 2 allows for justified interference with the right where the interference is prescribed by law, in pursuit of one of the legitimate aims, such as national security and the protection of the reputation or rights of others, and necessary in a democratic society.

When examining whether US acts and proposals are able to be transposed to address the anti-information issue, their compliance with the ECHR is necessary. Domestically, the Human Rights Act

[37] Frederick Schauer, 'The Second-Best First Amendment' (1989) 31(1) William & Mary Law Review 1-23, 5.
[38] Barendt (n 37) 13, Schauer, 'The Second-Best First Amendment' (n 38) 5, and Fish (n 36) 123. See, for example, Thomas Scanlon, 'A Theory of Freedom of Expression' (1972) 1(2) Philosophy & Public Affairs 204-226.
[39] Barendt (n 37) 18, Schauer, 'The Second-Best First Amendment' (n 38) 5, and Fish (n 36) 123. See, for example, Alexander Meiklejohn, Free Speech and Its Relation to Self-Government (Harper & Brothers 1948).
[40] Barendt (n 37) 21, Schauer, 'The Second-Best First Amendment' (n 38) 5, and Fish (n 36) 123.
[41] Barendt (n 37) 21-2.

1998 requires that judges interpret legislation consistently with the ECHR,[42] and if it is not possible to do so, to issue a declaration of incompatibility.[43] A declaration of incompatibility does not affect the validity of the legislation.[44] Whilst legislation that is not compliant may not trigger any legal effect domestically,[45] failure to comply may lead to a finding that the State has violated rights that are protected under the Convention. Therefore, Convention compliance, particularly compliance with Art.10, should be examined.

The First Amendment to the US Constitution is the first of ten amendments that make up the Bill of Rights. The right it protects is a negative freedom – 'Congress shall make no law... abridging the freedom of speech...' There is no provision for interference with the right; textually, the right appears to be absolute. However, and despite judicial assertions to the contrary,[46] there is a limit to the protection it affords.[47]

Given that discussion of legislative acts and proposals from the US is to serve as inspiration for the UK, discussions of their compliance with the First Amendment will act as an "alert" when considering possible use of a similar instrument in the UK. Issues that arise under the First Amendment may be faced also under Art.10. As there is no discussion over what approach to the anti-information issue is "best," there is no risk of asserting either free speech norm over the other.

[42] Human Rights Act 1998, s3.
[43] ibid s4.
[44] ibid s4(6).
[45] There is an incidental legal effect of a declaration of incompatibility which allows for a minister to amend incompatible legislation under a "fast track" procedure – see s10, Human Rights Act 1998.
[46] *Konigsberg v State Bar of California* [1961] 366 US 36, 61 (United States of America).
[47] For example, "obscene" speech – *Roth v US* [1957] 354 US 476 (United States of America).

Chapter 2: The Anti-Information Issue

This chapter investigates the nature of the anti-information issue. Firstly, it shall do so from the perspective of the truth-seeking theories, which will be followed with a discussion of the inapplicability of those theories to the anti-information issue. This chapter concludes that employing a traditional, "more speech," approach to minimising belief in false information alone is insufficient.

This chapter consists of three parts. Part 1 discusses the argument from truth and the marketplace of ideas. It demonstrates that those arguments concern debatable "truths" as opposed to verifiable facts. In doing so, it brings the applicability of the truth-seeking theories to the anti-information issue into question.

Parts 2 and 3 then discuss what renders the truth-seeking theories inapplicable to the anti-information issue. Part 2 distinguishes the rational person that is assumed and required by the truth-seeking theories from the idea-consumer, the ordinary person who comes to believe an idea that they encounter in the marketplace of ideas. It does so by reference to psychological research on the biases that underpin decision-making. Part 2 additionally informs the analysis of the likely effectiveness of measures discussed in Chapters 4, 5 and 6.

Part 3 is about the role technology plays in the dissemination of anti-information. It discusses the impact technology has had on the dissemination of false information, and in particular, targeted advertisements and bots. This part serves two purposes: to highlight the difficulties in countering anti-information in the modern day, and; to demonstrate the relevance of discussing statutes and proposals that address technology being used for the dissemination of anti-information in Chapters 5 and 6.

Part 1 concludes that the argument from truth and marketplace of ideas do not concern verifiable facts and preliminarily questions their applicability to a world where matters of factual truth form a part of other debates. Then Part 2 demonstrates that the idea-consumer is not rational and as such, concludes that it cannot be assumed that discussion, or "more speech," is capable of producing truth. Finally, Part 3 concludes that the difficulties that have arisen in countering anti-information due to technological

development make solutions to the anti-information issue that are compliant with traditional interpretations of the truth-seeking theories seem absurd.

Chapter 2.1: Truth-Seeking Speech Theories

This part will explain the truth-seeking theories in more detail. It is necessary to understand their origins in order to understand their scope and subsequently their applicability to the anti-information issue. By identifying that they are intended to address debatable ideas as opposed to verifiable facts, a lack of applicability to the anti-information issue becomes evident.

Chapter 2.1.1: The Argument from Truth

The argument from truth contends that freedom of expression, and the discussion it facilitates, is necessary because it allows for the discovery of the truth. Its origin is found in a number of works of political philosophy. John Milton, in 1644, wrote 'Let her [Truth] and Falsehood grapple; who ever knew Truth put to the worse, in a free and open encounter?'[1] Milton was in reality not concerned with the 'legal regulation of factual falsity' and was actually protesting his non-receipt of a publication license for an essay of his.[2]

John Stuart Mill, in 1859, published *On Liberty* 'the most extensive and important discussion'[3] of the argument from truth. Mill's argument was formulated on the notion of the "assumption of infallibility"[4] that speech must not be suppressed because if it is true, we lose an opportunity to replace error for truth.[5] He extends this to partially true speech, which in its collision with other partially true speech, allows for the remaining truth to be discovered.[6] Further, speech must not be suppressed because if it is false, an opportunity to refute it and gain a 'livelier impression of truth' is missed.[7] Without a truth being

[1] John Milton, *Areopagitica with a Commentary by Sir Richard C. Jebb and With Supplementary Material* (1918 Cambridge University Press) 58.
[2] Frederick Schauer, 'Facts and the First Amendment' (2010) 57 UCLA Law Review 897-919, 903.
[3] ibid 904.
[4] John Stuart Mill, *On Liberty* (Yale University Press 2003) 88.
[5] ibid 87.
[6] ibid 118.
[7] ibid 87.

'frequently' defended, Mill feared it becoming a 'dead dogma' – that is, a truth that is "assented to undoubtingly" for which a 'tenable defence' could not be given by the believer.[8]

Mill's argument however, like Milton's, does not concern verifiable fact. A verifiable fact cannot be "partially true." Neither is it something upon which one "exercises their intellect and judgment" because it is something that 'is considered necessary for [them] to hold opinions on'[9] like 'morals, religion, politics, social relations and the business of life.'[10] Verifiable facts were never meant to be included in his argument.[11] Rather, it was intended for 'values and feelings about facts'[12] as is evident in the examples of topics for debate Mill used,[13] including 'the morality of [certain doctrines]'[14] and 'the belief in a God.'[15]

Chapter 2.1.2: The Marketplace of Ideas

The "marketplace of ideas" is an evolutionary strand of the argument from truth that exists within the US Supreme Court's jurisprudence on the First Amendment to the US Constitution which states that no law shall be made 'abridging the freedom of speech.' It originates in the dictum of Justice Holmes in *Abrams v US* in which he argues that the 'best test of truth is the power of the thought to get itself accepted in the competition of the market.'[16] Holmes' dictum has been referred to as merely an appealing articulation of the argument from truth, as shorthand for the argument generally.[17] As the jurisprudence of the First Amendment developed, the argument evolved into its own legal doctrine.

The origin of the marketplace of ideas in the argument from truth is evident in *Whitney v California,* where Justice Brandeis explains that to avert 'falsehood and fallacies... the remedy to be applied is more

[8] ibid 103.
[9] ibid.
[10] ibid 104.
[11] Ari Ezra Waldman, 'The Marketplace of Fake News' (2018) 20 University of Pennsylvania Journal of Constitutional Law 845-70, 869, and Schauer (n 2) 905.
[12] Waldman (n 11) 868.
[13] Schauer (n 2) 905.
[14] Mill (n 4) the New Testament at 114, Pagan and Christian ethics at 115, and, public responsibility and private life at 116.
[15] Mill (n 4) 93.
[16] [1919] 250 US 616, 630.
[17] Gregory Brazeal, 'How much does a belief cost?: Revisiting the Marketplace of Ideas' (2011) 21(1) Southern California Interdisciplinary Law Journal 1-46, 5.

speech, not enforced silence.'[18] This became known as the counterspeech doctrine in First Amendment jurisprudence. He claims that the drafters of the First Amendment believed that the freedoms "to think and to speak" are 'indispensable to the discovery and spread of political truth.'[19] Justice Holmes concurs with the opinion given in *Whitney*,[20] and it is evident that like Milton and Mill, the advocates of the marketplace theory believed freedom of expression was for the discovery of truth on debatable topics, as opposed to verifiable facts.

Even in *Brandenburg v Ohio*,[21] which overturned *Whitney*, the same attitude is expressed. The Supreme Court gives examples of a number of types of "speech" including picketing,[22] the burning of a draft card,[23] and goes as far as to say there is no constitutional difference between 'advocacy of abstract ideas' and 'advocacy of political action.'[24] For the most part, the jurisprudence never considers verifiable facts – 'Demonstrably false statements were not part of this jurisprudence, or of its intellectual tradition. The marketplace of ideas was always meant to be a marketplace of *ideas*, not *facts*.'[25] At least, not until *New York Times v Sullivan*,[26] which found that the First Amendment's protection applied to some false statements in the context of the law of defamation.[27] Even after this point, how the First Amendment was to treat 'issues of verifiable fact or demonstrable factual falsity' may have been "touched upon," but it was certainly never the focus of the free speech tradition, in the US or abroad, until the latter part of the twentieth century.[28]

Brandeis' opinion in *Whitney* included an important qualification of "more speech, not enforced silence" in that it was the remedy to be applied 'If there be time to expose through discussion the falsehood and

[18] [1927] 274 US 357, 377.
[19] ibid 375.
[20] ibid 380.
[21] [1969] 395 US 444.
[22] ibid 455.
[23] ibid 456.
[24] ibid 457.
[25] Waldman (n 11) 869.
[26] *New York Times v Sullivan* [1964] 376 US 254.
[27] Defamatory statements against public officials would lose First Amendment protection if made with 'actual malice' (*New York Times* (n 26) 280). '[E]rroneous statement is inevitable in free debate, and that it must be protected if the freedoms of expression are to have the "breathing space" that they "need . . . to survive."' (at 271-2).
[28] Schauer (n 2) 906-7.

fallacies...'[29] The counterspeech doctrine has been cited since *Whitney*,[30] despite it being overturned, particularly in recent years, reaffirming the US Supreme Court's commitment to the doctrine.[31] The counterspeech doctrine, and the argument from truth generally, rely upon time which makes them ineffective against certain types of speech.

The jurisprudence has reflected this since its beginnings, in *Schenck v US*, it was said that speech creating a "clear and present danger" would not be protected, such as someone 'falsely shouting fire in a theatre and causing a panic.'[32] Later in *Brandenburg v Ohio*, the "clear and present danger" test was replaced by a question of whether speech 'incit[es] or produc[es] imminent lawless action.'[33] A broader but similar standard to that used originally in *Whitney* – 'imminent danger.'[34] However, speech that creates an "imminent danger" or "incites lawless action" is not the only speech where there is insufficient time to correct "falsehoods and fallacies" through "more speech."

Not all anti-information creates an imminent danger yet, there is not always sufficient time to correct its harms, particularly with strategic disinformation. Coordinated disinformation campaigns leave insufficient time for the falsehoods and fallacies they propagate to be corrected. Facebook said that Russian operatives made 80,000 posts to sway US politics which were seen by 126 million Americans from 2015 to 2017,[35] and it had found Russian accounts that bought $100,000 in advertisements for the 2016 Presidential election.[36] The 20 most popular stories from "hoax sites and hyperpartisan blogs" generated 8,711,000 interactions on Facebook, making them more popular that real, mainstream news in the build-up to the election.[37] The popularity of that disinformation, particularly that it outperformed

[29] *Whitney* (n 18) 377.
[30] *Linmark Associates, Inc. v Township of Willingboro* [1977] 431 US 85.
[31] For example, see *Lorillard Tobacco Co v Reilly* [2001] 533 US 525, 586, and *US v Alvarez* [2012] 567 US 1, 7 (Justice Kennedy).
[32] [1919] 249 US 47, 52.
[33] *Brandenburg* (n 21) 447.
[34] *Whitney* (n 18) 373.
[35] David Ingram, 'Facebook says 126 million Americans may have seen Russia-linked political posts' (Reuters, 30 October 2017) <https://www.reuters.com/article/us-usa-trump-russia-socialmedia/facebook-says-126-million-americans-may-have-seen-russia-linked-political-posts-idUSKBN1CZ2OI> accessed 22 April 2022.
[36] Scott Shane and Vindu Goel, 'Fake Russian Facebook Accounts Bought $100,000 in Political Ads' *The New York Times* (New York City, 6 September 2017) <https://www.nytimes.com/2017/09/06/technology/facebook-russian-political-ads.html> accessed 22 April 2022.
[37] Craig Silverman, 'This Analysis Shows How Viral Fake Elections News Stories Outperformed Real News on Facebook' (Buzzfeed News, 16 November 2016) <https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook> accessed 22 April 2022.

"real" news just before the election, demonstrates the ineffectiveness of the marketplace of ideas to counteract it in sufficient time. Regardless of how well those stories performed, the fact that they were not identified as "fake" until after the election shows that the marketplace of ideas is too slow in correcting certain falsehoods and fallacies. Given that an election can be won by the smallest change in voter preferences, any harm that may have been done could be highly influential.[38]

The development of the marketplace jurisprudence demonstrates the implicit assumption that the argument for truth 'was as applicable to factual' as to other "truths"[39] and the application of this jurisprudence to the issue of anti-information demonstrates that the reliance on counterspeech is ineffective due to its requirement of "sufficient time" which is not always available.

In *US v Alvarez*, Justice Kennedy said, 'The remedy for speech that is false is speech that is true. This is the ordinary course in a free society. The response to the unreasoned is the rational; to the uninformed, the enlightened; to the straight-out lie, the simple truth.'[40] He staunchly backs counterspeech in the *ordinary* course of *free* society. He leaves unanswered questions such as: when is the unordinary course required, and; is a society free when the information upon which it governs itself is so easily manipulated?

Chapter 2.1.3: Debatable Ideas versus Verifiable Facts

The argument from truth and the marketplace of ideas do not concern "factual" truths and as such, the arguments fail to account for a world where debate concerns matters of factual truth as a part of debates on other subjects. Since statements of facts can be verified by reference to standards applicable to all, unlike opinions which are determined solely be reference to the speaker, an attempt to 'penalise false statements of fact [is] in theory consistent with a position of neutrality'[41] and therefore, it is consistent

[38] For example, for the 2016 Presidential Election, only 12% of third-party voters would have needed to have been persuaded by Russian interference to vote for Trump over Clinton to have won the election for Trump – Jane Mayer, 'How Russia Helped Swing the Election for Trump' *The New Yorker* (New York City, 24 September 2018) <https://www.newyorker.com/magazine/2018/10/01/how-russia-helped-to-swing-the-election-for-trump> accessed 22 April 2022.
[39] Schauer (n 2) 907.
[40] *Alvarez* (n 31) 15-6.
[41] Robert Post, 'The Constitutional Concept of Public Discourse: Outrageous Opinion, Democratic Deliberation, and Hustler Magazine v. Falwell' (1990) 103(3) Harvard Law Review 601-86, 660.

with the argument from truth's principle of infallibility and the principles of the marketplace of ideas.[42] '[S]tatements of fact are not arguments, and the very ability to argue presupposes accurate facts.'[43]

However, such an argument may raise questions regarding what a verifiable truth is and both who is able and who should be given the power to verify "truths." The answers to such questions must avoid censoring speakers,[44] particularly those expressing minority viewpoints,[45] but must remain effective. The exact answers may differ from one state to another that seeks to punish false expression, given their unique conceptions of democracy and their particular political, historical and cultural circumstances.[46]

Nonetheless given that penalisation of such speech would be a severe response, to avoid any chilling effect, it could be limited to expressions that cause a particular harm or are in pursuit of a particular gain, as with fraud.[47] Additionally, any enforcement should come from a regulator independent of the State in order to avoid allegations of it being a "Ministry of Truth,"[48] and its methodology for finding a fact to be false should be transparent.[49]

Ultimately, freedom of expression 'is a farce unless factual information is guaranteed and the facts themselves are not in dispute... factual truth informs political [ideas].'[50] Regardless, the arguments'

[42] Content-neutrality is to be discussed further in Chapter 4 onwards.
[43] Post (n 41).
[44] See for example, Amnesty International, 'A Human Rights Approach to Tackle Disinformation: Submission to the Office of the High Commissioner for Human Rights' (14 April 2022) 12-3 <https://www.amnesty.org/en/wp-content/uploads/2022/04/IOR4054862022ENGLISH.pdf> accessed 17 April 2023.
[45] '[D]emocracy does not simply mean that the views of a majority must always prevail' – *Young, James and Webster v UK* App. No.7601/76, [63].
[46] *Hirst v UK* App. No.74025/01, [61].
[47] Fraud Act 2006, s2(1)(b) cf s5(2)-(4).
[48] See Alex Green, 'Mark Steyn show on GB News breached Ofcom code with Covid claims' (The Independent, 6 March 2023) <https://www.independent.co.uk/news/uk/ofcom-gb-news-naomi-wolf-b2294926.html> accessed 1 June 2023, and Mark Steyn, 'The Show Ofcom Won't Let You See' (Steyn Online, 16 March 2023) <https://www.steynonline.com/13331/the-show-ofcom-wont-let-you-see> accessed 1 June 2023.
See also, George Trefgarne, 'The Spirit of Orwell lives – that's the Ministry of Truth of it' *The Telegraph* (London, 16 July 2001) <https://www.telegraph.co.uk/finance/2726243/The-spirit-of-Orwell-lives-thats-the-Ministry-of-Truth-of-it.html> accessed 1 June 2023.
Also, Margi Murphy, 'U.S. Plan to Track Misinformation Sparks Its Own Misinformation' (Bloomberg, 11 May 2022) accessed 1 June 2023.
[49] A transparent methodology could help to avoid censorship allegations by limiting its findings of falsity to statistical fallacies such as cherry-picked statistics – for an infamous example, '£350 million EU claim "a clear misuse of official statistics"' <https://fullfact.org/europe/350-million-week-boris-johnson-statistics-authority-misuse/> accessed 17 April 2023.
[50] Hannah Arendt, *Between Past and Future* (Viking Press 1968) 238, in Post (n 41) 660.

reliance on time and more truthful speech as the remedy to anti-information are ineffective and inadequately address the threat posed by anti-information on democracy.

Chapter 2.2: The Inapplicability of a "More Speech" Approach to the Anti-Information Issue

This section discusses what difficulties may be encountered when attempting to counter the effects of anti-information. Anti-information is made difficult to counter because of many psychological tendencies that interfere with the information selection, belief formation and evaluative stages of decision-making.

The remedy desired by the argument from truth and the marketplace of ideas for falsity is discussion of an opposing viewpoint – "more speech." However, it is rendered weak due to the decision-making biases and the role technology plays in the dissemination of anti-information. The truth-seeking arguments assume that the idea-consumer rationally comes to their beliefs, but that is not the case.

This section will distinguish the economically rational consumer in the ordinary market from the idea-consumer in the marketplace of ideas. The first part will briefly explain what economic rationality is and why it is not realistic. Then, how biases interfere with the information the idea-consumer bases their beliefs upon will be explained to separate the idea-consumer from the rational consumer at the point of information selection. It will then be shown that the idea-consumer does not come to believe an idea in the same way that an economically rational actor "chooses" a product. Finally, biases that interfere with how idea-consumers evaluate ideas they encounter in the marketplace of ideas will be used to separate the idea-consumer from the rational consumer at the evaluative stage of decision-making.

Chapter 2.2.1: Economic Rationality

The marketplace of ideas metaphor has been taken literally by some, and economic ideas have been imposed upon it.[51] As scholars considered the implications of behavioural economics on the marketplace of ideas model, it faced further criticism,[52] reflecting the general criticism economic rationality has faced, that it is not realistic.[53]

An economically rational actor makes decisions that are 'the best means to the chooser's ends... [achieving their goal] with the greatest margin of benefit over cost.'[54] Arguably, for an idea-consumer, this would mean choosing ideas that make them 'feel good-that is, [ideas that] give [them] the most utility.'[55] Although maximising the amount of truth is most often perceived to be the aim of the marketplace of ideas,[56] that does not necessarily mean the truth will offer the most utility to idea-consumer – 'it is better to be the contented fool than the unhappy Socrates.'[57] Generally though, the economically rational idea-consumer would choose true ideas over false ones.[58]

[51] Richard Posner, *Economic Analysis of Law* (7th ed, Aspen Publishers 2007) 727.
[52] See for example Derek Bambauer, 'Shopping Badly: Cognitive Biases, Communications, and the Fallacy of the Marketplace of Ideas' (2006) 77(3) University of Colorado Law Review 649-710.
See also, Cass R. Sunstein, *Behavioral Law and Economics* (2012 Cambridge University Press).
[53] Amartya Sen, 'Rational Fools: A Critique of the Behavioral Foundations of Economic Theory' (1977) 6(4) Philosophy & Public Affairs 317-344, 317-22.
[54] Richard Posner, 'Rational Choice, Behavioural Economics, and the Law' (1997) 50 Stanford Law Review 1551, 1551.
[55] Darren Bush, 'The Marketplace of Ideas: Is Judge Posner Chasing Don Quixote's Windmills?' (2000) 32(4) Arizona State Law Journal 1107-48, 1112.
[56] Brazeal (n 17) 27.
[57] Bush (n 55) 1112.
[58] Brazeal (n 17) 20. Additionally, it has been demonstrated that people prefer to share accurate information – see, Ziv Epstein, Adam J. Berinsky, Rocky Cole, Andrew Gully, Gordon Pennycook, and David G. Rand, 'Developing an accuracy-prompt toolkit to reduce COVID-19 misinformation online' (2021) 2(3) Harvard Kennedy School Misinformation Review <https://misinforeview.hks.harvard.edu/article/developing-an-accuracy-prompt-toolkit-to-reduce-covid-19-misinformation-online/> accessed 10 April 2023 , and Gordon Pennycook, Ziv Epstein, Mohsen Mosleh, Antonio A. Arechar, Dean Eckles & David G. Rand, 'Shifting attention to accuracy can reduce misinformation online' (2021) 592(7855) Nature, available at <https://www.nature.com/articles/s41586-021-03344-2> accessed 10 April 2023.

Rationality is a high standard of decision-making to assume upon an individual and there is little evidence to accept it as an assumption.[59] Rationality is attractive for its simplicity,[60] but it is not very realistic. As Brietzke argues,

> '[I]t ignores a host of factors that make us human, including altruism, habit, bigotry, panic, genius, luck or its absence, and factors such as peer pressures, institutions, and cultures that turn us into social animals. A dehumanized, desocialized, and often sexist "economic man" supposedly goes through life as if it were one long series of analogies to isolated transactions on the New York Stock Exchange.'[61]

The economically rational actor can be distinguished from the idea-consumer in a number of ways. Common biases lead the information which the idea-consumer selects to base their decision upon to be reduced to that which gives them reason to be optimistic, which serves their interests, and which confirms the beliefs they already hold. Further, the idea-consumer does not "choose" what to belief in the same way the economically rational actor chooses the action which brings them the most utility. Finally, whilst "choosing" what to believe, similar biases to those encountered at the information selection stage causes the idea-consumer to evaluate information in a manner that is not economically rational.

Chapter 2.2.2: Information Selection

At the information selection stage, a number of biases may detract rationality from the idea-consumer's decision-making ability. Biases at this stage often result in the idea-consumer relying upon anti-information over verifiable facts in forming their beliefs.

[59] Christopher Wonnell, 'Truth and the Marketplace of Ideas' (1986) 19(3) UC Davis Law Review 669-728, 673, and Paul Brietzke, 'How and Why the Marketplace of Ideas Fails' (1997) 31(3) Valparaiso University Law Review 951-969, 953-7.
[60] Raymond Boudon, 'Limitations of Rational Choice Theory' (1998) 104(3) American Journal of Sociology 817-28, 17.
[61] Paul H. Brietzke, 'Urban Development and Human Development' (1992) 25(3) Indiana Law Review 741-98, 753.

Optimism bias causes people to ignore pertinent information. It occurs in roughly 80% of the population,[62] and is 'the tendency to assess a lower probability for oneself to experience negative health events compared to others.'[63] It can lead people to not read warnings on products,[64] to have lower concern for the future consequences of climate change,[65] and to undermine preparedness due to unrealistic expectations about the future.[66] It especially occurs in relation to perceptions of new risks.[67] This is particularly relevant for health-related anti-information, as we often consume information about health risks. As a result, the idea-consumer 'may underconsume, or ignore, information on salient topics because [they] think [they] are safer than [they] actually are.'[68]

A self-serving bias describes situations where a person emphasises or gives priority to information that strengthens the position they are defending. As a result, people can rely upon different information to justify differing conclusions even when provided with the same totality of information, and when holding the same collective aim.[69] Additionally, 'we perceive our evaluations as impartial and disinterested, but suspect others of succumbing to self-interest.'[70] Although it is not inconceivable that rational idea-consumers could reach different conclusions,[71] where a self-serving bias is present they cannot rationally have reached opposing viewpoints. Although disagreement over ideas is inevitable in

[62] Tali Sharot, 'The optimism bias' (2011) 21(23) Current Biology R941-R945, R942.
[63] Elena Druică, Fabio Musso and Rodica Ianole-Călin, 'Optimism Bias during the Covid-19 Pandemic: Empirical Evidence from Romania and Italy' (2020) 11(3) Games 1 <https://www.mdpi.com/2073-4336/11/3/39> accessed 22 April 2023.
[64] Barbara Luppi & Francesco Parisi, 'Beyond Liability: Correcting Optimism Bias through Tort Law' (2009) 35(1) Queen's Law Journal 47-66, 51.
[65] Sabine Pahl, Stephen Sheppard, Christine Boomsma, and Christopher Groves, 'Perceptions of time in relation to climate change' (2014) 5(3) WIREs Climate Change 375-88, 379.
[66] Geoffrey Beattie, 'Optimism bias and climate change' (2018) 33 British Academy Review 12-15, 15.
[67] Joan Costa-Font, Elias Mossialos, and Caroline Rudisill, 'Optimism and the perceptions of new risks' (2009) 12(1) Journal of Risk Research 27-41, 38.
[68] Derek Bambauer (n 52) 676.
[69] ibid 677, cf Kimberly A Wade-Benzoni, Ann E Tenbrunsel and Max H Bazerman, 'Egocentric Interpretations of Fairness in Asymmetric, Environmental Social Dilemmas: Explaining Harvesting Behavior and the Role of Communication' (1996) 67(2) Organizational Behavior and Human Decision Processes 111-26, 125 and Linda Babcock and George Loewenstein, 'Explaining Bargaining Impasse: The Role of Self-Serving Biases' (1997) 11(1) Journal of Economic Perspectives 109-26, 112-115.
[70] Bambauer (n 52) 678, cf Wade-Benzoni et al. (n 69) 125.
[71] 'There is no more reason to assume that an ideally efficient idea-market is one in which all consumers arrive at the same conclusions than there is to assume that an ideally efficient appliance-market is one in which all consumers possess toasters of the same size and colour.' Brazeal (n 17) 29.

the truth-seeking justification of freedom of expression, it is evident that self-serving biases may prevent a truth being discovered that otherwise would be. Thus, in this manner, the idea-consumer is not rational.

Confirmation bias occurs where an individual searches, interprets or recalls information consistently with their previously held beliefs so as to not prejudice those views, but rather to support or confirm those beliefs.[72] Further, confirmation bias is stronger when presented with information that is 'attitude-consistent' – aligning with beliefs already held.[73] This phenomenon has been observed for political views.[74] Whilst searching and recalling information relates to the idea-consumer's selection of information upon which to base their view, the effect of confirmation bias on their interpretation of information means that it affects the evaluative stage of decision-making also.

The idea-consumer is unable to rationally form beliefs (consume ideas) in the marketplace of ideas because the information they select to make their decision is incomplete, flawed and biased. The idea-consumer omits information that does not align with the belief they hold by recalling only that which confirms their belief, by giving greater priority to information that confirms their belief, and in general by neglecting information that supports attitude-inconsistent ideas.

Chapter 2.2.3: "Choosing" in the Marketplace of Ideas

One key distinction between the rational actor and the idea-consumer is that a rational actor in a market is said to *choose* between available options, which is generally not the case for idea-consumers.[75]

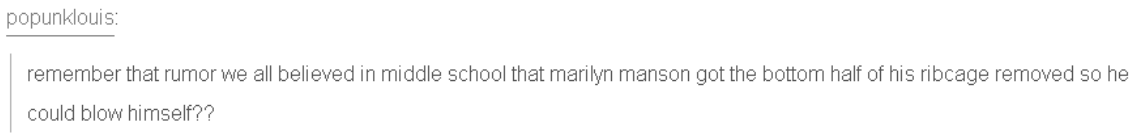
The economically rational actor thoughtfully makes their decisions, whilst the same cannot be said of the idea-consumer – someone that consumes many ideas passively, whose knowledge can be displaced by repetition of information, and is likely to have many unconscious beliefs.

[72] Raymond Nickerson, 'Confirmation Bias: A Ubiquitous Phenomenon in Many Guises' (1998) 2(2) Review of General Psychology 175-220, 175.
Yanmengqian Zhou and Lijiang Shen, 'Confirmation Bias and the Persistence of Misinformation on Climate Change' (2021) Communications Research 1-24, 3-4.
[73] Zhou and Shen (n 72) 5.
[74] In a study, the authors examined political views regarding the Iraq War, tax cuts and stem cell research. Brendan Nyhan and Jason Reifler, 'When Corrections Fail: The Persistence of Political Misperceptions' (2010) 32(2) Political Behaviour 303-330.
[75] ibid 7 and 15.

Epidemic Belief

Whilst the economically rational actor is an active participant in the decision-making process, the idea-consumer is generally not actively involved in their obtaining of beliefs. False ideas may spread similarly to an epidemic where the idea-consumer attains beliefs passively, without critical deliberation.[76] Rumours and "hoaxes" demonstrate how ideas spread this way.

Stories of celebrities and wealthy individuals having had their ribs removed, often for vanity,[77] are fairly common. This tumblr post has nearly 300,000 "notes" (total interactions of users liking, sharing and commenting anything that originates from the original post).[78]



Whilst such rumours are almost certainly myths,[79] that has not prevented them being repeated with minor details changed, and with little discussion of their veracity. Whether it be Cher,[80] Usher,[81] or 'ultra-fashionable Victorians,'[82] and whether the purpose was vanity or autofellatio, the idea that a person had their ribs removed for a non-medical reason persisted, and adapted to the times – becoming connected with the most popular people of the time.[83]

This persistent idea, whether true or false, has been described as a "meme" – a unit of cultural information – which can be used to model cultural evolution, as originally posited by Richard

[76] Brazeal (n 17) 17.
[77] Barbara Mikkelson & David Mikkelson, 'Did Cher Have Ribs Removed To Make Her Waist Smaller?' (Snopes, 22 June 2000) <https://www.snopes.com/fact-check/getting-waisted/> accessed 13 June 2023.
[78] The original account has been deactivated but the post is still visible having been reshared to other public accounts and can be found with a Google search <https://www.google.com/search?q=site%3Atumblr.com+marilyn+manson+got+the+bottom+half+of+his+ribcage+removed> accessed 13 June 2023.
[79] Mikkelson & Mikkelson (n 77). See also, Valerie Steele, *The Corset – A Cultural History* (Yale University Press 2003) 2 and 72-4.
[80] Mikkelson & Mikkelson (n 77).
[81] See an updated version of that post from tumblr user 'powerburial' <https://powerburial.tumblr.com/post/88245733235/popunklouis-remember-that-rumor-we-all-believed> accessed 13 June 2023.
[82] Steele (n 79) 73.
[83] Mikkelson & Mikkelson (n 77). Recently it has been connected to Kim Kardashian, see Alanna McKnight, 'The Kurious Kase of Kim Kardashian's Korset' (2020) 3(1) Fashion Studies <https://www.fashionstudies.ca/the-kurious-kase-of-kim-kardashians-korset> accessed 13 June 2023.

Dawkins.[84] Although the memetic model for explaining the spread of cultural information has been highly criticised, it proves to be a good model that demonstrates how ideas passively spread in the marketplace of ideas.

Criticism of the model often argues that the original conception of memes describes them as being "self-replicating" which is said to be antithetical to culture as a *product* of human agency.[85] However, regarding the spread of ideas online, a popular definition accounts for the human role in producing internet memes requiring that they are 'created with awareness of each other; and were circulated, imitated and/or transformed via the Internet by many users.'[86]

Nonetheless, determining whether memetics explains the human role in producing information is not necessary for it to be an accurate model of how beliefs are attained by the idea-consumer. A person is likely to reproduce beliefs they hold, and that is relevant to explaining how ideas spread. However, in distinguishing the rational agent and the idea-consumer, what ideas they believe and why they believe them is more revealing than their role as a disseminator.[87] The passive, memetic model of information transmission still adequately describes idea-consumption.

Memetic logic has been used to examine the spread of anti-information online. Dawkins' "meme" and "memetics" have been used to describe the spread of anti-information online,[88] with the term "meme" being used mostly interchangeably with its common meaning – 'An image, video, piece of text, etc.,

[84] See Richard Dawkins, *The Selfish Gene: 40th Anniversary Edition* (2016 OUP) 245-260.
[85] Henry Jenkins, Sam Ford and Joshua Green, *Spreadable Media* (2013 New York University Press) 19.
[86] Limor Shifman, *Memes in Digital Culture* (2014 MIT Press) 41. See, for example, Michael Johann and Lars Bülow, 'One Does Not Simply Create a Meme: Conditions for the Diffusion of Internet Memes' (2019) 13 International Journal of Communication 1720-1742, 1723; Pratiwi Utami, 'Hoax in Modern Politics: The Meaning of Hoax in Indonesian Politics and Democracy' (2018) 22(2) Jurnal Ilmu Sosial Dan Ilmu Politik 85-97, 89, and; Mike Hajimichael, 'Social Memes and Depictions of Refugees in the EU: Challenging Irrationality and Misinformation with Media Literacy Intervention' in Alison MacKenzie, Jennifer Rose and Ibrar Bhatt, *The Epistemology of Deceit in a Postdigital Era: Dupery by Design* (2021 Springer) 205.
[87] '[A] supply-side model of the marketplace of ideas ... could lead us to believe a market was functioning optimally simply because producers' incentives were optimised—even if the true ideas they produced found no adherents, and even if false ideas... received widespread acceptance.' – Brazeal (n 17) 11. The issue with that perspective is that measuring the marketplace of ideas from the perspective of the speaker omits the consumption of ideas entirely, therefore any "measurement" cannot be used to evaluate the idea market for its ability to maximise truth or minimise factual falsity.
[88] See for example, Utami (n 86), Hajimichael (n 86).

typically humorous in nature, that is copied and spread rapidly by internet users, often with slight variations.'[89]

Rodríguez-Ferrándiz et al. examined certain "hoaxes" (pieces of disinformation) and their capacity to become memetic,[90] particularly because of the role of the internet in increasing the potential of memes and anti-information to propagate.[91] They distinguished the virality of a post and its capacity to become memetic, where virality is the popularity of a particular piece of information and where memetic describes the processes in which information is repeated and often transformed.[92] A singular piece of information can be viral but "mutates" into a meme when it becomes the thing from which many pieces of information have evolved.[93] Like the idea of a person having their ribs removed for non-medical reasons.

Their study includes a story that claimed dolphins had been caught on video swimming in a marina due to inactivity of the port because of COVID lockdowns,[94] which was refuted for four different ports in Spain,[95] but also in Ukraine, Italy, Taiwan, Portugal, Mexico and India.[96] Whilst the rib removal rumour

[89] 'meme, n.' (*OED Online*, OUP April 2023) <https://www.oed.com/view/Entry/239909> accessed 8 April 2023. See, for example, Andrew Moshirnia, 'Who Will Check the Checkers? False Factcheckers and Memetic Misinformation' (2020) 4 Utah Law Review 1029-74.

[90] Raúl Rodríguez-Ferrándiz, Cande Sánchez-Olmos, Tatiana Hidalgo-Marí and Estela Saquete-Boro, 'Memetics of Deception: Spreading Local Meme Hoaxes during COVID-19 1st Year' (2021) 13(6) Future Internet <https://www.mdpi.com/1999-5903/13/6/152> accessed 22 April 2023. Whilst they distinguish the types of hoaxes, they do not clearly define hoaxes themselves. They expand upon previous research which had defined hoaxes as 'all the false content that is disseminated to the public, intentionally manufactured for multiple reasons, that can range from a simple joke or parody, to an ideological argument, including for economic fraud.' (translation by author, Ramón Salaverría, Nataly Buslón, Fernando López-Pan, Bienvenido León, Ignacio López-Goñi, and María-Carmen Erviti, 'Desinformación en tiempos de pandemia: tipología de los bulos sobre la Covid-19' (2020) 29(3) Profesional de la información 1-17, 4.) Given this definition, "hoaxes" can be replaced with "disinformation."

[91] Rodríguez-Ferrándiz et al. (n 90) 4-5, cf 'we live in an era driven by hypermemetic logic, in which almost every major political event sprouts a stream of memes' Shifman (n 86) 4.

[92] Rodríguez-Ferrándiz et al. (n 90) 5.

[93] ibid cf Shifman (n 86) 56-57.

[94] Rodríguez-Ferrándiz et al. (n 90) 14-15.

[95] 'FALSE: A video where dolphins appear to be swimming in a marina, supposedly due to the inactivity of the port caused by the coronavirus health crisis. This video has been circulated saying that it is the Promenade of Palma de Mallorca, the port of Denia (Alicante), the port of Moraira (Alicante) or the port of Premià de Mar (Barcelona).' (Poynter., 18 April 2020) <https://www.poynter.org/?ifcn_misinformation=a-video-where-dolphins-appear-swimming-in-a-marina-supposedly-due-to-the-inactivity-of-the-port-caused-by-the-coronavirus-health-crisis-this-video-has-been-circulated-saying-that-it-is-the-promenade> accessed 13 June 2023.

[96] Rodríguez-Ferrándiz et al. (n 90) 14-5.

adapted to the time in which it was disseminated, the idea of dolphins swimming in unused ports was adapted for the locality in which it was shared.

By examining hoaxes/pieces of disinformation which started in a particular locality and became more widespread, they concluded that they were 'opportunistic, adaptable [and] highly contagious.'[97] This epidemic spread of anti-information makes it difficult to be refuted in a way anticipated by the truth-seeking theories. The "infection" should be defeated by being disproven through discussion, however, because of anti-information's "significant qualities" of fecundity and ability to mutate,[98] there is no single vaccination, or rebuttal, that inoculates the idea-consumer to an anti-information meme which spreads widely. Refuting one claim that dolphins did not swim through a local port might not prevent an idea-consumer believing that dolphins did not swim through any port whilst it was closed during a COVID lockdown. Similarly, Cher proving she has all her ribs,[99] has not prevented the rumour arising for others.

Rodríguez-Ferrándiz et al. conclude that anti-information can be 'analysed as a memetic practice.'[100] The hoaxes/disinformation they examined were not "tested by the market," they were reproduced elsewhere without the truthfulness being verified.[101] Anti-information may spread passively without critical deliberation,[102] and can mutate like a virus, adapting to circumstances such as the time and locality of its dissemination. Therefore, the rational actor can be distinguished from the idea-consumer because of the idea-consumer's susceptibility to this fecund and adaptable form of information.

[97] ibid 17.
[98] ibid 16.
[99] Mikkelson & Mikkelson (n 77).
[100] Rodríguez-Ferrándiz et al. (n 90) 17.
[101] Rodríguez-Ferrándiz et al. (n 90).
[102] Brazeal (n 17) 17.

Idea Repetition and Belief

The idea-consumer can also be distinguished from the rational actor where information is repeated to them. The illusory truth effect – the notion that the more you encounter certain information, the more likely you are to believe it – has been well observed.[103]

It is reasonable for a person to rely on how familiar information feels (which increases the more they encounter it) given it 'often proves to be an accurate and cognitively inexpensive strategy,'[104] as opposed to "searching for stored knowledge,"[105] meaning to test what information they encounter against their knowledge. Therefore, knowledge does not protect someone against their believing information that is repeated to them.[106] The idea-consumer is not rationally assessing every piece of information they encounter, using all their knowledge, and it would be unreasonable to expect that of them.

Given that anti-information spreads analogously to an epidemic,[107] an idea-consumer may passively encounter anti-information multiple times, even if it is not repeatedly directed towards them.[108] Even simple exposure 'to false information will increase belief in the false information.'[109] Inoculating someone against perceiving repeated information as more truthful is not as simple as providing a warning. Doing so has been found to have no significant effect as,[110] despite increasing general

[103] The original study concluded that 'the repetition of a plausible statement increases a person's belief in the referential validity or truth of that statement.' – Lynn Hasher, David Goldstein and Thomas Toppino, 'Frequency and the Conference of Referential Validity' (1977) 16(1) Journal of Verbal Learning and Verbal Behavior 107-112, 111.
[104] Lisa K. Fazio, Nadia M. Brashier, B. Keith Payne and Elizabeth J. Marsh, 'Knowledge Does Not Protect Against Illusory Truth' (2015) 144(5) Journal of Experimental Psychology 993-1002, 1000.
[105] ibid.
[106] ibid 999-1000.
[107] The illusory truth effect supports this also. Information appears more accurate *because* it is repeated – see, Valentina Vellani, Sarah Zheng, Dilay Ercelik, Tali Sharot, 'The illusory truth effect leads to the spread of misinformation' (2023) 236 Cognition 6, <https://www.sciencedirect.com/science/article/pii/S0010027723000550> accessed 22 April 2023., 'If repeated exposure biases people to share news more, the longer information circulates, the higher the probability that it will be considered as true and further shared with others' at 6.
[108] Directed at them, such as in an election campaign, or courtroom – Danielle C. Polage, 'Making up History: False Memories of Fake News Stories' (2012) 8(2) Europe's Journal of Psychology 245-250, 249.
[109] Polage (n 108) 248. Also, Gordon Pennycook, Tyrone D. Cannon and David G. Rand, 'Prior exposure increases perceived accuracy of fake news' (2018) 147(12) Journal of Experimental Psychology 1865-80, 1874.
[110] Pennycook et al., 'Prior exposure increases perceived accuracy of fake news' (n 109) 1874.

scepticism and willingness to share,[111] the perceived accuracy of the information grows regardless and the positive effects of the warning are cancelled out.[112]

Additionally, the more familiar the information seems, a person is not only more likely to believe it but also more likely to attribute it to a credible source,[113] which further solidifies the belief. This effect is stronger when the information that is encountered is consistent with beliefs the person already holds,[114] and can lead to the spontaneous generation of a credible source when recalling the information, potentially leading to information being "upgraded" in believability as it passes from one person to another.[115] Though it is still present even when the repeated information is inconsistent with held beliefs.[116]

Unlike the rational actor, making fiscal decisions that bring them the most utility, the idea-consumer can be made to believe an idea by having it repeated to them.

Unconscious Belief

A final significant way the idea-consumer can be differentiated from the economically rational actor is through the role unconscious beliefs have upon their decision-making process, which may prevent "rational" choice.

There is significant psychological evidence to support the existence of unconscious or "implicit," biases, "attitudes" or beliefs.[117] Implicit feelings[118] and past experience[119] have been shown to bias the decision-making process – the 'unconscious operation of stereotypes'[120] demonstrates this.

[111] ibid 1876.
[112] ibid.
[113] Alison R. Fragale and Chip Heath, 'Evolving Informational Credentials: The (Mis)Attribution of Believable Facts to Credible Sources' (2004) 30(2) Personality and Social Psychology Bulletin 225-236, 227-31, and; Polage (n 108) 248.
[114] Fragale and Heath (n 113) 231-33.
[115] ibid 234.
[116] Pennycook et al., 'Prior exposure increases perceived accuracy of fake news' (n 110).
[117] Regarding their unconscious nature and the scope of their unconscious features, see Bertram Gawronski, Wilhelm Hofmann and Christopher J. Wilbur, 'Are "implicit" attitudes unconscious?' (2006) 15(3) Consciousness and Cognition 485-99.
[118] R. J. Dolan, 'Emotion, Cognition and Behaviour' (2002) 298 Science 1191-4, 1194.
[119] Anthony G. Greenwald & Mahzarin R. Banaji, 'Implicit social cognition: Attitudes, self-esteem, and stereotypes' (1995) 102(1) Psychological Review 4-27, 4-5.
[120] ibid 15.

'Implicit attitudes are introspectively unidentified (or inaccurately identified) traces of past experience that mediate favourable or unfavourable feeling, thought, or action toward social objects.'[121] The majority of people have implicit attitudes,[122] and their impact has been observed outside of a controlled environment – in the real world.[123]

Arguably, explicit bias, that being bias known to the decision-maker, could be accounted for and the idea-consumer need not depart from the economically rational actor. There is potential for them to exclude explicit bias from their decision, and therefore still reach the conclusion that holds the greatest utility. However, the existence of implicit biases separates the rational actor from the idea-consumer.

Although they are malleable,[124] the introspectively unidentifiable nature of implicit bias prevents the idea-consumer from being able to definitively exclude an implicit bias from the decision-making process. Individuals cannot tell if they have an implicit bias through self-reflection. Individuals tend to be unsettled upon discovering they have implicit biases that conflict with their explicit beliefs,[125] yet they may through training unlearn such biases.[126] Though, like a rubber band, they will likely snap back

[121] ibid 8.
[122] Anthony G. Greenwald and Lina Hamilton Krieger, 'Implicit Bias: Scientific Foundations' (2006) 94(4) California Law Review 945-967, 955-58, and Brian A. Nosek, Frederick L. Smyth, Jeffrey J. Hansen, Thierry Devos, Nicole M. Lindner, Kate A. Ranganath, Colin Tucker Smith, Kristina R. Olson, Dolly Chugh, Anthony G. Greenwald and Mahzarin R. Banaji, 'Pervasiveness and correlates of implicit attitudes and stereotypes' (2007) 18 European Review of Social Psychology 36-88, 75.
[123] Dan-Olof Rooth, 'Implicit Discrimination in Hiring: Real World Evidence' (2007) IZA Discussion Paper 2764, and Nilanjana Dasgupta, 'Implicit Ingroup Favoritism, Outgroup Favoritism, and Their Behavioral Manifestations' (2004) 17(2) Social Justice Research 143-69.
[124] See Nilanjana Dasgupta, 'Implicit Attitudes and Beliefs Adapt to Situations: A Decade of Research on the Malleability of Implicit Prejudice, Stereotypes and the Self-Concept' in Patricia Devine, Ashby Plant (eds., Vol 47), *Advances in Experimental Social Psychology* (2013 Academic Press) 233-79. See also, for examples of implicit biases being altered, Nilanjana Dasgupta and Anthony G. Greenwald, 'On the Malleability of Automatic Attitudes: Combating Automatic Prejudice with Images of Admired and Disliked Individuals' (2001) 81(5) Journal of Personality and Social Psychology 800-14, and Patricia G. Devine, 'Stereotypes and Prejudice: Their Automatic and Controlled Components' (1989) 56(1) Journal of Personality and Social Psychology 5-18.
[125] Ditte Marie Munch-Jurisic, 'The Right to Feel Comfortable: Implicit Bias and the Moral Potential of Discomfort' (2020) 23(1) Ethical Theory and Moral Practice 237-250, 238.
[126] For example, through direct exposure to counterstereotypical persons, vicarious exposure to counterstereotypes, increasing motivation to be egalitarian, education (Jerry Kang, Mark Bennett, Devon Carbado, Pam Casey, Nilanjana Dasgupta, David Faigman, Rachel Godsil, Anthony G. Greenwald, Justin Levinson, Jennifer Mnookin, 'Implicit Bias in the Courtroom' (2012) 59(5) UCLA Law Review 1124-1886, 1169, 1170-1, 1174, 1174-5 and 1181-4).

without repeated training.[127] Having to maintain such an arduous process in order to exclude implicit biases would suggest the idea-consumer is not able to act as an economically rational actor.

The ways we develop implicit attitudes bears particular relevance to the anti-information issue. The general explanation is that they develop throughout an individual's lifetime beginning at a young age,[128] as they are exposed to different messages.[129] One source that causes implicit attitudes to develop is media and the news.[130]

One model for explaining how we come to believe an idea is the "automatic update" model, where 'we believe a proposition in the very act of considering it, and that once considered, it is only through conscious scrutiny that we "unbelieve" it.'[131] Strange qualifies this model by arguing that the automatic beliefs of "fictional worlds" do not apply to the "actual world,"[132] only to 'the worlds of dreams, wishes, stories and the like.'[133] In a world where it is increasingly difficult to determine what is the "fictional" and what is the "actual" world, the possibility of our beliefs being affected by anti-information is very real. Some observations have already been made on the effects anti-information can have on our unconscious minds. Anti-information can covertly manipulate our behaviour,[134] without an individual being aware of the existence of the manipulation, its source, its content, or its impact on their behaviour.[135]

---

[127] Mahzarin R. Banaji and Anthony G. Greenwald, 'Blindspot: Hidden Biases of Good People' (2013 Delacorte Press) 152.
[128] Laurie A. Rudman, 'Sources of Implicit Attitudes' (2004) 13(2) Current Directions in Psychological Science 79-82, 79-80, and Luigi Castelli, Cristina Zogmaister and Silvia Tomelleri, 'The Transmission of Racial Attitudes Within the Family' (2009) 45(2) Developmental Psychology 586-91, 589-90.
[129] Jerry Kang, 'Bits of Bias' in Justin D. Levinson and Robert J. Smith (eds.), *Communications Law* (2012 Cambridge University Press) 132-145, 132-139, and Dasgupta (n 124), 237.
[130] Kang (n 129) 135-139 and Dasgupta (n 124).
[131] Jeffrey J. Strange, 'How Fictional Tales Wag Real-World Beliefs' in Melanie C. Green, Jeffrey J. Strange and Timothy C. Brock, *Narrative Impact: Social and Cognitive Foundations* (2002 Taylor & Francis Group) 263-86, 274. cf on the automatic update model: Daniel T. Gilbert, 'How Mental Systems Believe' (1991) 46(2) American Psychologist 107-119, and Richard J. Gerrig, *Experiencing narrative worlds: On the psychological activities of reading* (1993 Yale University Press).
[132] Strange (n 131) 275.
[133] ibid.
[134] Zach Bastick, 'Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation' (2021) 116 Computers in Human Behavior <https://www.sciencedirect.com/science/article/pii/S0747563220303800> accessed 13 June 2023.
[135] ibid 7.

The economically rational actor unlike the idea-consumer thoughtfully makes their decisions. The idea-consumer instead tends to obtain ideas passively – as they would an infection – by encountering the idea, particularly on repeated occasions. Additionally, they hold unconscious beliefs – implicit attitudes developed subconsciously over a lifetime due things they have experienced.

Chapter 2.2.4: Biases in the Evaluative Stage of Decision-Making

Similarly, biases at the evaluative stage of decision-making detract from the idea-consumer's rational decision-making ability.[136] These biases can cause the idea-consumer to determine their beliefs irrationally by causing them to value certain information more favourably than they should. They also tend to support the status quo.

People tend to have an aversion to loss. 'They are more displeased with losses than they are pleased with equivalent gains.'[137] Even with favourable odds, we fear this risk of loss more than we value the opportunity to gain.[138] This subsequently causes individuals to have a tendency to prefer the status quo, 'because the disadvantages of leaving it loom larger than advantages.'[139] Similarly, individuals tend to prefer certain outcomes over uncertain outcomes.[140]

Confirmation bias also results in sources that present attitude-consistent information as being perceived as having greater expertise and being more trustworthy than sources that challenge held beliefs by presenting attitude-inconsistent information.[141] An individual is less likely to find an expert "trustworthy and knowledgeable" when they adopt an attitude-inconsistent opinion, and their perceived trustworthiness has greater sway than their perceived expertise.[142]

[136] Bambauer (n 52) 693.
[137] Cass R. Sunstein, *Behavioral Law and Economics* (2012 Cambridge University Press) 5.
[138] Daniel Kahneman, Jack L. Knetsch, and Richard H. Thaler, 'The Endowment Effect, Loss Aversion and Status Quo Bias' (1991) 5(1) Journal of Economic Perspectives 193-206, 199-203.
[139] ibid 197-8.
[140] Amos Tversky and Daniel Kahneman, 'The Framing of Decisions and the Psychology of Choice' (1981) 211(4481) Science 453-458, 455-6.
[141] Zhou and Shen (n 72) 17, and Rebecca Helm and Hitoshi Nasu, 'Regulatory Responses to 'Fake News' and Freedom of Expression: Normative and Empirical Evaluation' (2021) 21(2) Human Rights Law Review 302-28, 306.
[142] Helm and Nasu (n 141) 306.

Anti-intellectualism 'has been growing in the mass public for decades.'[143] The increase in this attitude makes it increasingly difficult to counter anti-information with expert knowledge and opinion. Anti-intellectualism has been 'significantly associated with decreased belief in expert consensus with respect to climate change and nuclear power safety'[144] and 'plays a key role in explaining why individuals support expert-averse movements and political candidates'[145] including Donald Trump and Brexit.[146]

Similarly, implicit biases lead to snap judgements and interfere with rational decision-making where the idea-consumer does not commit to deliberating the decision, as opposed to making a decision spontaneously, or without particular deliberation.[147]

A few egocentric biases are particularly relevant when considering the capacity truth has for countering anti-information.

Firstly, the third person effect finds 'that people will tend to overestimate the influence that mass communications have on the attitudes and behaviour of others… [and] will expect the communication to have a greater effect on others than on themselves.'[148] This effect has been observed online,[149] and

[143] Matthew Motta, 'The Dynamics and Political Implications of Anti-Intellectualism in the United States' (2018) 46(3) American Politics Research 465-498, 466. In the UK, for example, Michael Gove saying "People in this country have had enough of experts." See David Matthews, 'Brexit would be a victory for those who distrust academics' <https://www.timeshighereducation.com/blog/brexit-would-be-victory-those-who-distrust-academics> accessed 22 April 2022, ibid 493.
See also, Josh Lowe, 'Michael Gove: I'm "Glad" Economic Bodies Don't Back Brexit' <https://www.newsweek.com/michael-gove-sky-news-brexit-economics-imf-466365> accessed 22 April 2022, ibid 493.
[144] Motta (n 143) 480.
[145] ibid 481.
[146] ibid 482. This finding regards American support for Brexit.
[147] John Dovidio, Kerry Kawakami and Samuel Gaertner, 'Implicit and Explicit Prejudice and Interracial Interaction' (2002) 82(1) Journal of Personality and Social Psychology 62-68, 66, and Jens Asendorpf, Rainer Banse, and Daniel Mücke, 'Double Dissociation Between Implicit and Explicit Personality Self-Concept: The Case of Shy Behavior' (2002) 83(2) Journal of Personality and Social Psychology 380-393, 391.
[148] W Phillips Davison, 'The Third-Person Effect in Communication' (1983) 47(1) Public Opinion Quarterly 1-15, 3.
[149] Nikos Antonopoulos, Andreas Veglis, Antonis Gardikiotis, Rigas Kotsakis, and George Kalliris, 'Web Third-person effect in structural aspects of the information on media websites' (2015) 44 Computers in Human Behavior 48-58, 54-6.

with regards to anti-information in particular,[150] where the "others" that were assumed to be susceptible to anti-information were generally political "others."[151]

The effect is linked to "information overload" where people feel as though they cannot process all the information they encounter,[152] particularly because with the flood of news 'users might begin to skim content without verifying its accuracy and authenticity.'[153] Therefore, it would be worsened during political and major cultural events, when news media increases.

If the idea-consumer believes they are unlikely to be affected by anti-information, then they are more susceptible to it and less likely to take action against it. This consumer is not able to reason as rationally as assumed, though they may think they are able to.

Another egocentric bias relevant to the anti-information issue is the false consensus effect, which causes individuals 'to see their own behavioural choices and judgments as relatively common and appropriate to existing circumstances while viewing alternative responses as uncommon, deviant, or inappropriate.'[154]

Since social media encourages interaction with likeminded individuals, 'heavier social media users exhibit higher rates of false consensus effects.'[155] The effect occurs through 'exposure to a biased news feed, even [without] further interaction.'[156] This bias has been observed in online communities of neo-Nazis and environmentalists.[157] Conservative individuals may be more susceptible to the effect, tending

[150] S Mo Jang and Joon K Kim, 'Third person effects of fake news: Fake news regulation and media literacy interventions' (2018) 80 Computers in Human Behavior 295-302, 299, and Nicoleta Corbu, Denisa-Adriana Oprea, Elena Negrea-Busuioc and Loredana Radu, '"They can't fool me, but they can fool the others!" Third person effect and fake news detection' (2020) 35(2) European Journal of Communication 165-180, 174-5.
[151] Jang and Kim (n 150), and Corbu et al. (n 150) 175-6.
[152] Shuo Tang, Lars Willnat and Hongzhong Zhang, 'Fake news, information overload and the third-person effect in China' (2021) 6(4) Global Media and China 492-507, 495-6.
[153] ibid 496.
[154] Lee Ross, David Greene and Pamela House, 'The "false consensus effect": An egocentric bias in social perception and attribution processes' (1977) 13(3) Journal of Experimental Social Psychology 279-301, 280.
[155] Cameron Bunker and Michael Varnum, 'How strong is the association between social media use and false consensus?' (2021) 125 Computers in Human Behavior 2 and 5 <https://www.sciencedirect.com/science/article/pii/S0747563221002703> accessed 22 April 2023.
[156] Robert Luzsa & Susanne Mayr, 'False Consensus in the Echo Chamber: Exposure to Favorably Biased Social Media News Feeds Leads to Increased Perception of Public Support for Own Opinions' (2021) 15(1) Cyberpsychology: Journal of Psychosocial Research on Cyberspace <https://cyberpsychology.eu/article/view/12254> accessed 22 April 2023.
[157] Magdalena Wojcieszak, 'False Consensus Goes Online: Impact of Ideologically Homogeneous Groups on False Consensus' (2008) 72(4) Public Opinion Quarterly 781-91, 788.

to overestimate the extent to which others share their beliefs, whereas liberals tend to underestimate and fail to capitalise on support they do have.[158]

Neo-Nazis were not protected against the effect by encountering dissimilar opinions offline and may have been "immunised" to counterspeech.[159] However, the effect may be counteracted through exposure to counterspeech online, that being 'dissimilar news diets,'[160] and user interest/involvement in the topic.[161]

The false consensus effect means that it is more difficult for counterspeech to replace beliefs based upon anti-information because the idea-consumer's false beliefs are validated by their belief they are in the majority, and they see the counterspeech as "uncommon, deviant or inappropriate." The idea-consumer therefore cannot act as the rational actor that prefers truth.

Naïve realism is related to false consensus and may be an explanation for it.[162] The idea-consumer is separated from the rational actor in a broader manner according to naïve realism, where the individual believes that their knowledge of the real world is accurate and their perception of it is faithful/objective, when it 'is at best indirect and mediated.'[163] It is broader than false consensus, rather than holding a belief they are in the majority, the individual fails to appreciate they see facts through the lens of their

[158] Chadly Stern, Tessa West and Peter Schmitt, 'The Liberal Illusion of Uniqueness' (2014) 25(1) Psychological Science 137-144, 142. Though this was only partially replicated for views on vaccination which would suggest it depends on the particular content of the message, a conclusion which is supported elsewhere (Luzsa & Mayr (n 156). Conservatives overestimated the extent other conservatives agreed with them on vaccination, underestimated liberals and had no distorted views compared to the general population, in a sample of mostly pro-vaccination subjects - Mitchell Rabinowitz, Lauren Latella, Chadly Stern and John Jost, 'Beliefs about Childhood Vaccination in the United States: Political Ideology, False Consensus, and the Illusion of Uniqueness' (2016) 11(7) PLoS ONE <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0158382> accessed 13 June 2023.
[159] Magdalena Wojcieszak, 'Computer-Mediated False Consensus: Radical Online Groups, Social Networks and News Media'(2011) 14(4) Mass Communication and Society 527-546, 540.
[160] ibid.
[161] Luzsa & Mayr (n 156). The authors also mention ambiguity tolerance which is a psychological trait describing the comfort of a person in face of contradictory information.
[162] Lee Ross, Mark Lepper and Andrew Ward, 'History of Social Psychology: Insights, Challenges, and Contributions to Theory and Application' in Susan Fiske, Daniel Gilbert, and Gardner Lindzey, *Handbook of Social Psychology* (Vol.1, 5th edn Wiley 2010) 3-50, 23.
[163] ibid 22.

own experiences, knowledge, and feelings.[164] Naïve realism is considered "foundational" to social psychology.[165]

Naïve realism has effects beyond the false consensus effect that separate the idea-consumer from the rational actor. Principally, it causes a bias blind spot where we do not recognise bias in ourselves but do so in others,[166] this can cause 'Misunderstanding, mistrust,... and unwarranted pessism about the ability to find common ground with those with whom we disagree.'[167] The bias blind spot causes the idea-consumer to not value the ability of discussion to find truth as well as to cause the misunderstanding and mistrust which makes it harder for discussion to find commonality.

Reactive devaluation is an effect caused by naïve realism where proposals or concessions from an adversary have diminished attractiveness versus that from a non-adversary.[168] Such effect is exploitable, particularly in politics, where through calling an alternative unreasonable, a speaker may gain the sympathy of previously uninterested parties.[169] The effect is worsened 'When pessimism is widespread and rooted in a history of past negotiation failures.'[170] Therefore, the longer dispute as to the truth exists and the more anti-information polarises the debate, the less likely the idea-consumer is capable of a "meaningful compromise,"[171] and the less likely a "truth" can be agreed upon.

Finally, speakers act as "idea-rating agents" determining the value the idea-consumer places in a particular idea for them.[172] An idea-consumer that places trust in a politician,[173] social media site,[174] and

[164] ibid 23.
[165] ibid 22.
[166] Emily Pronin, Daniel Lin, Lee Ross, 'The Bias Blind Spot: Perceptions of Bias in Self Versus Others' (2002) 28(3) Personality and Social Psychology Bulletin 369-381, 378.
[167] ibid 379.
[168] Lee Ross and Constance Stillinger, 'Barriers to Conflict Resolution' (1991) 7(4) Negotiation Journal 389-404, cf Lee Ross and Andrew Ward, 'Naïve Realism in Everyday Life: Implications for Social Conflict and Misunderstanding' in Edward Reed, Elliot Turiel and Terrance Brown (eds), *Values and Knowledge* (Psychology Press 1996) 103-135, 126-7.
[169] Lee Ross, 'Reactive Devaluation in Negotiation and Conflict Resolution' in Kenneth Arrow, Robert Mnookin, Lee Ross, Amos Tversky, and Robert Wilson, *Barriers to Conflict Resolution* (W W Norton & Company 2007) 26-42, 28.
[170] ibid 42.
[171] ibid.
[172] Brazeal (n 17) 33.
[173] ibid. Also, Simon Gächter and Elke Renner, 'Leaders as role models and "belief managers" in social dilemmas' (2018) 154 Journal of Economic Behavior & Organization 321-334, 331.
[174] Xizhu Xiao, Porismita Borah and Yan Su, 'The dangers of blind trust: Examining the interplay among social media news use, misinformation identification, and news trust on conspiracy beliefs' (2021) 30(8) Public Understanding of Science 977-992, 986.

family and friends,[175] is more likely to believe an idea they disseminate or their rating of an idea because 'people tend to seek accurate information from sources they consider reliable and to avoid the exposure to sources they distrust.'[176] Influential individuals exercise force disproportionate to the "truth" of their ideas on the marketplace of ideas.

Anti-information takes advantage of a number of biases that affect the evaluative stage of decision-making. Firstly, anti-information that supports the status quo has an advantage due to the aversion to loss and confirmation biases which cause an idea-consumer to reinforce already held beliefs. Implicit biases cause the idea-consumer to make snap judgments without testing their thinking. Egocentric biases and naïve realism cause the idea-consumer to think that mass communications do not affect their opinions, to think they are within the public consensus, that their perception of the world is objective or at least faithful, to miss biases they have because of a "blind spot" in their view of themselves, and to place less value in ideas that come from "adversaries" such as those with differing political opinions.

Chapter 2.2.5: Biases, Idea-Consumption and the G.I. Joe Consumer

The section has demonstrated in three ways the idea-consumer can be distinguished from the economically rational actor. Principally, the idea-consumer does not rationally select the information upon which they determine their beliefs. The information the idea-consumer relies upon is incomplete and biased. It is motivated by psychological tendencies that favour optimism, information that serves the idea-consumers interests and confirmation of beliefs they already hold.

Additionally, the idea-consumer does not "choose" their beliefs in the marketplace of ideas like the economically rational actor chooses a product in the market. Their consumption of ideas and formation of beliefs more resembles an epidemic as information is passively received. Further, they develop

[175] Brazeal (n 17) 33. Also, Daniel Bar-Tal, Alona Raviv and Tali Freund, 'An Anatomy of Political Beliefs: A Study of Their Centrality, Confidence, Contents, and Epistemic Authority' (1994) 24(10) Journal of Applied Social Psychology 849-872.
[176] Alberto Ardèvol-Abreu and Homero Gil de Zúñiga, 'Effects of Editorial Media Bias Perception and Media Trust on the Use of Traditional, Citizen, and Social Media News' (2017) 94(3) Journalism & Mass Communication Quarterly 703-724, 706.

beliefs without realising, through repeated exposure to the idea. Experiences also causes them to generalise, and form attitudes unconsciously.

Ultimately, were the idea-consumer able to select information rationally and actively choose what to believe, they would likely choose poorly because the idea-consumer does not rationally evaluate the ideas they encounter. The idea-consumer is opposed to change, preferring to avoid the risk of loss that may come by believing something new. They believe they are objective and unaffected by anti-information, or any mass communication. They believe they are in the majority and that their opposition is irrational. Hence they struggle to find agreement ("truth"), not trusting their opposition, believing them to be biased, irrational and untrustworthy.

The idea-consumer is unable to fulfil the truth-seeking role envisioned by the truth-seeking theories because of the difficulty they face in selecting the information upon which they form their beliefs, the way they form beliefs, and the difficulty they face in evaluating the new ideas they encounter. Particularly, a finding of truth through the combination of partial truths, as described by Mill,[177] is made difficult due to polarisation of debate where the idea-consumer believes their opposition to be biased, irrational and untrustworthy.

A worthy disclaimer at this point is that knowledge of these psychological tendencies and biases is insufficient to avert them. That belief is itself a "misguided notion" which some have called the G.I. Joe fallacy, after a 1980s television series called "G.I. Joe" which had the closing tagline, 'Now you know. And knowing is half the battle.'[178] Though the notion that knowledge is sufficient to avert bias has been disproven elsewhere.[179] The idea-consumer, the author, and presumably the reader should recognise that they are subject to these biases. However, this section does not serve as a call for "rational

[177] Mill (n 4) 118.
[178] Ariella Kristal and Laurie Santos, 'G.I. Joe Phenomena: Understanding the Limits of Metacognitive Awareness on Debiasing' (2021) Harvard Business School Working Paper 21-084, 3 <https://www.hbs.edu/faculty/Pages/item.aspx?num=59722> accessed 14 June 2023.
[179] For example, Fazio et al. (n 104) 999-1000, and Ross et al. (n 162) 22. See also, Robert Vallone, Lee Ross and Mark Lepper, 'The Hostile Media Phenomenon: Biased Perception and Perceptions of Media Bias in Coverage of the Beirut Massacre' (1985) 49(3) Journal of Personality and Social Psychology 577-585.

debate" and is simply a recognition that discussion is not inherently rational, nor can it be assumed that discussion, overseen and conducted by the irrational idea-consumer, is capable of producing truth.

Chapter 2.3: The Role of Technology in the Dissemination of Anti-Information

Amongst the younger generation, social media is replacing traditional news sources.[180] Whilst outright lies and propaganda have been used for centuries to persuade and mislead, the difference now is the speed and volume at which such ideas can be spread.[181] Anti-information can diffuse 'farther, faster, deeper and more broadly than the truth in all categories of information.'[182]

Knowledge of how the marketplace of ideas differs online, compared to traditional media, is necessary to understand the issue. As such, this section will outline how incentives for disseminating ideas and the degree of gatekeeping in speech has changed, as well as the role played by targeted advertisements, algorithms and bots. This will demonstrate how technological developments interfere with the role discussion/counterspeech supposedly plays in the discovery of the truth, and show that time as a remedy to falsity cannot be relied upon. Given "recent" technological developments (compared to the broadcast era), there is insufficient time to remedy anti-information through discussion alone.

[180] Ofcom, 'News Consumption in the UK: 2022' (21 July 2022) 2, available at <https://www.ofcom.org.uk/__data/assets/pdf_file/0024/241827/News-Consumption-in-the-UK-Overview-of-findings-2022.pdf> accessed 20 April 2022.
[181] Anya Schiffrin, 'Disinformation and Democracy: The Internet Transformed Protest but Did Not Improve Democracy' (2017) 71(1) Journal of International Affairs 117-126, 121.
[182] Soroush Vosoughi, Deb Roy, and Sinan Aral, 'The Spread of True and False News Online' (2018) Science 1146, 1147.

Chapter 2.3.1: Technological Development and the Online Marketplace of Ideas

This part will discuss how the economic incentive for producing ideas has changed compared to the broadcast era and how technological development, and the subsequent absence of gatekeeping, has led to a lower quality of speech being encountered.

The Changing Economic Incentive of Idea Dissemination

Technological advancement has "globalised" media, changing how 'information is captured, shared and discussed around the world.'[183] In particular, the accessibility of technology for publishing websites and earning money from advertisements on such sites has incentivised the creation of stories that will be shared widely.[184] Speakers and the platforms they publish on can make money with every "click,"[185] and "fake news" gets more clicks.[186] With the increasing accessibility of the internet and growing potential audience,[187] the economic incentive for sharing and allowing anti-information to be shared grows. As long as the reduction of anti-information remains antithetical to the commercial interests of speakers and platforms, we are unlikely to see voluntary action taken by those parties.[188]

Parasitic journalism is an example of 'a thriving business model' [189] which the current media environment has encouraged. Parasitic journalism is low cost for the speaker as they 'recycle and

[183] Matthew Davis and Per Fors, 'Towards a Typology of Intentionally Inaccurate Representations of Reality in Media Content' (2020) 590 IFIP Advances in Information and Communication Technology 291-304, 292.
[184] Peter Fernandez, 'The technology behind fake news' (2017) 34(7) Library Hi Tech News 1-5, 3.
[185] For example, a Washington Post interview with someone that 'make[s] like $10,000 a month from AdSense [Google's advertising program that allows users to monetise their websites].' – Caitlin Dewey, 'Facebook fake-news writer: "I think Donald Trump is in the White House because of me"' (2016 Washington Post) <https://www.washingtonpost.com/news/the-intersect/wp/2016/11/17/facebook-fake-news-writer-i-think-donald-trump-is-in-the-white-house-because-of-me/> accessed 12 April 2023.
[186] Richard Spearman, 'Fake news and financial market blues' (2017) 8 Journal of International Banking and Financial Law 488-90, 489.
[187] Internet users made up 7% of the world population in 2000 and 60% in 2020 – The World Bank, 'Individual using the Internet (% of population)' <https://data.worldbank.org/indicator/it.net.user.zs> accessed 23 April 2022.
[188] Spearman (n 186). Also, Samuel Rhodes, 'Filter Bubbles, Echo Chambers, and Fake News: How Social Media Conditions Individuals to Be Less Critical of Political Misinformation' (2022) 39(1) Political Communication 1-22, 15.
[189] Phillip Napoli, 'What If More Speech Is No Longer the Solution? First Amendment Theory Meets Fake News and the Filter Bubble' (2018) 70 Federal Communications Law Journal 55-104, 69.

recirculate' a story rather than producing a costly original story.[190] A radio broadcast of *The War of the Worlds* caused mass hysteria because of news programs greatly exaggerating a story about a few people fleeing their homes believing it to be true.[191] This model has been employed by the disseminators of anti-information also.[192]

Other than the changing economic incentive, the absence of journalistic gatekeeping online has changed the dynamics of the marketplace of ideas.

The Absence of Gatekeeping

More people are using online sources to get their news, which do not 'adhere to typical standards of truth, scientific inquiry, and evidence-based news and information' and which detract from usage of traditional sources,[193] which previously gatekept news production, upholding those standards.[194] Lack of gatekeeping online, where speakers can communicate directly to large audiences, has allowed those 'that would otherwise not have the resources to conduct disinformation campaigns with traditional mass media [the opportunity to do so online].'[195]

In the past, gatekeeping improved the quality of ideas by preventing false ideas reaching idea-consumers.[196] This was because there was an incentive to appear neutral and objective – when there were fewer speakers, attaining and retaining a large audience meant quality journalism was preferred.[197] Gatekeeping meant that any "fake news" came from the same source as news, however, the internet, by

[190] ibid 70.
[191] Daniela Manzi, 'Managing the Misinformation Marketplace: The First Amendment and the Fight Against Fake News' (2019) 87(6) Fordham Law Review 2623-2651, 2624.
[192] Napoli (n 189) 70 cf Craig Silverman & Lawrence Alexander, 'How Teens in The Balkans Are Duping Trump Supporters with Fake News' (Buzzfeed, 3 November 2016) <https://www.buzzfeednews.com/article/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo> accessed 19 April 2023.
[193] Schiffrin (n 181) 123.
[194] Napoli (n 189) 71-74.
[195] PR Chamberlain, 'Twitter as a Vector for Disinformation' (2010) 9(1) Journal of Information Warfare 11, 11.
[196] Napoli (n 189) 71-2.
[197] ibid 72.

giving more people the capacity to speak, has increased the numbers of disseminators of both information and anti-information.[198]

Parasitic journalism, and a lack of gatekeeping, means anti-information can proliferate incredibly quick. A man with 40 Twitter followers 'became part of a national controversy' after wrongfully claiming anti-Trump protests had been manufactured.[199]

Lack of gatekeeping can also be exploited by those with resources to distributed propaganda in a clandestine manner. State-sponsored troll farms can inundate parts of the internet with propaganda.

Internet "trolls" are people that behaviour 'in a deceptive, destructive or disruptive manner in a social setting on the Internet... exploiting "hot-button issues" to make users appear overly emotional or foolish in some manner.'[200] They make use of the anonymity and greater opportunities online to connect with others and 'pursue their personal brand of "self-expression,"'[201] characterised by antisocial behaviour.

Internet users are told to not engage with such users and to instead ignore them,[202] but this does not stop the troll's acts from being "felt."[203] Users still see the abuse. Similarly, the effects of state-sponsored trolls cannot be avoided by simply ignoring them. Trolls use popular topics and "trends" to spread propaganda to everyone on a particular platform.[204] Russian and Iranian trolls have hijacked popular topics,[205] including the Black Lives Matter movement to spread anti-information.[206]

---

[198] Raúl Rodríguez-Ferrándiz, Cande Sánchez-Olmos, Tatiana Hidalgo-Marí and Estela Saquete-Boro, 'Memetics of Deception: Spreading Local Meme Hoaxes during COVID-19 1st Year' (2021) 13(6) Future Internet 152, 2.
[199] 'Austin man says sorry for posting misleading anti-Trump protester Tweet' (13 November 2016, FOX 29 Philadelphia) <https://www.fox29.com/news/austin-man-says-sorry-for-posting-misleading-anti-trump-protester-tweet> accessed 20 April 2023.
[200] Erin Buckels, Paul Trapnell, and Delroy Paulhus, 'Trolls just want to have fun' (2014) 67 Personality and Individual Differences 97-102, 97.
[201] ibid 101.
[202] Wikitionary, 'do not feed the troll' <https://en.wiktionary.org/wiki/don%27t_feed_the_troll> accessed 21 April 2023.
[203] Erik Cambria, Praphul Chandra, Avinash Sharma, and Amir Hussain, 'Do Not Feel The Trolls' (2010) 664 CEUR Workshop Proceedings 1.
[204] Jarred Prier, 'Commanding the Trend: Social Media as Information Warfare' (2017) 11(4) Strategic Studies Quarterly 50-85, 59-60.
[205] Savvas Zannettou, Tristan Caulfield, William Setzer, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn, 'Who Let The Trolls Out? Towards Understanding State-Sponsored Trolls' (2019) WebSci '19: Proceedings of the 10th ACM Conference on Web Science 353-362, 361.
[206] Prier (n 199) 67-70.

Troll "factories" can inundate social media platforms through mass communication,[207] and when real users engage with this propaganda they lend it authenticity.[208] The characteristics and effects of trolls are similar to that of bots,[209] which are soon to be discussed. Bots can be utilised by trolls to increase their rate of communication.

Additionally, the availability of low-cost software for media production, whilst improving the ability of content creators to make entertainment online, has become a problem when combined with 'a business model that relies on attracting views and a general public's decreasing attention span...'[210] It is reasonable to anticipate a point in the future when due to these technologies, distinguishing visual media that represents real life and that which represents an inaccurate portrayal of real life will no longer be possible, like with written media.[211]

In summary, the incentives behind sharing ideas online have shifted since the broadcast era. High quality journalism is no longer required to gain more consumers. The economic incentives behind having more consumers means that the incentive to speak is driven by the number of attainable consumers. Rather than number of consumers influencing the incentive to create higher quality journalism, the number of consumers has become the priority.

Chapter 2.3.2: Personalised Information

Through targeted advertising, and other data-mining algorithms, intermediaries are able to limit what information an idea-consumer (e.g., social media user) is able to access, limiting the effects of

[207] Ulises Meijas and Nikolai Vokuev, 'Disinformation and the media: the case of Russia and Ukraine' (2017) 39(7) Media, Culture & Society 1027-42, 1034.
[208] ibid 1029.
[209] David A. Broniatowski, Amelia M. Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C. Quinn and Mark Dredze, 'Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate' (2018) 108(10) American Journal of Public Health 1378-1384.
Stefan Stieglitz, Florian Brachten, Björn Ross, and Anna Jung, 'Do Social Bots Dream of Electric Sheep? A Categorisation of Social Media Bot Accounts' (2017) ACIS 2017 Proceedings 89, 6.
[210] Davis and Fors (n 183) 293.
[211] ibid 296. For example, due to deepfake technology, see Mika Westerlund, 'The Emergence of Deepfake Technology: A Review' (2019) 9(11) Technology Innovation Management Review 40-53.

discussion by preventing the idea-consumer from being able to access sufficient information opposed to their viewpoint.[212]

Targeted Advertisements

Targeted advertisements are a result of technological advancement which makes anti-information so prolific. Micro-targeting is the practice of analysing personal data, collected online through interactions on social media,[213] to put people into groups which can then be shown advertisements tailored for them,[214] "better equipping" those with an interest in the dissemination of anti-information than in the past.[215] Importantly, although efforts were made in the past to use data to target people, the "technological capacity to target citizens has taken another leap forward" as demonstrated by the Cambridge Analytica scandal in 2016,[216] where Facebook user data was collected without their knowledge or consent in order to inform a political advertising system. People were categorised into groups, which were then targeted by advertisements, particularly the groups that were considered more susceptible. Cambridge Analytica micro-targeting is known to have been used in the campaigns of Ted Cruz and Donald Trump,[217] and potentially in the Leave.EU campaign for the 2016 Brexit Referendum,[218] although the official investigation by the Information Commissioner did not find any evidence of that being the case.[219]

Technology is now capable of much deeper profiling. Where before only broad inferences were able to be made through data such as magazine subscriptions and car purchases, now online activity reveals

[212] Napoli (n 189) 77-78.
[213] ibid 75.
[214] The Electoral Commission, 'Political Finance Regulation and Digital Campaigning: A Public Perspective : GfK UK report for qualitative research findings' 24 April 2018, 16.
[215] Napoli (n 189) 75.
[216] ibid 76.
[217] Matthew Kelly, 'Before Trump, Cambridge Analytica was on team Cruz' <https://www.opensecrets.org/news/2018/03/before-trump-cambridge-analytica-was-on-team-cruz/> accessed 22 April 2022.
[218] Reuters, 'What are the links between Cambridge Analytica and a Brexit campaign group?' <https://www.reuters.com/article/us-facebook-cambridge-analytica-leave-eu-idUSKBN1GX2IO> accessed 22 April 2022.
Alex Hern, 'Cambridge Analytica did work for Leave.EU, emails confirm' <https://www.theguardian.com/uk-news/2019/jul/30/cambridge-analytica-did-work-for-leave-eu-emails-confirm> accessed 22 April 2022.
[219] BBC News, 'Cambridge Analytica "not involved" in Brexit referendum, says watchdog' <https://www.bbc.co.uk/news/uk-politics-54457407> accessed 22 April 2022.

much more about the user and allows for individualised profiling.[220] The Leave Campaign of the 2016 Brexit Referendum sent approximately 'one billion targeted digital adverts' particularly near the postal voting deadline and the final 10 days of the campaign, which Dominic Cummings suggests is one of the main reasons their campaign was successful.[221]

Targeted advertisements have the potential to be extremely proficient for the dissemination of ideas and subsequently anti-information, particularly if used efficiently through deep profiling of the idea-consumer. Additionally, they leave the idea-consumer with little time to "test" the idea. Targeted advertisements can be used to overwhelm the idea-consumer with anti-information close to events such as elections. With little time for the anti-information to be refuted, the idea-consumer's capacity to refute it is ineffective. Should the idea-consumer act as a rational actor would with regards to a particular idea they encounter in a targeted advertisement, it may affect their beliefs before they have sufficient time to examine it.

Algorithms

An algorithm is 'a precisely defined set of mathematical or logical operations for the performance of a particular task [by a computer]'[222] Intermediaries such as social media sites use algorithms to keep users engaged with their platform so that the users encounter more advertisements and generate more revenue.[223] These algorithms analyse the data the social network accumulates about what a particular user interacts with and uses it to make predictions about what they might like, so that similar content can be shown to them.[224] Milan argues,

[220] Zeynep Tufekci, 'Engineering the public: big data, surveillance and computational politics' (2014) 19(7) First Monday <https://firstmonday.org/ojs/index.php/fm/article/view/4901/4097> accessed 13 June 2023.
[221] Dominic Cummings, 'On the referendum #20: the campaign, physics and data science – Vote Leave's 'Voter Intention Collection System' (VICS) now available for all' <https://dominiccummings.com/2016/10/29/on-the-referendum-20-the-campaign-physics-and-data-science-vote-leaves-voter-intention-collection-system-vics-now-available-for-all/> accessed 5 March 2022.
[222] Oxford English Dictionary, 'algorithm, n.' <https://www-oed-com.ezphost.dur.ac.uk/view/Entry/4959?redirectedFrom=algorithm#eid> accessed 21 April 2023.
[223] Sang Ah Kim, 'Social Media Algorithms: Why You See What You See' (2017) 2(1) Georgetown Law Technology Review 147-154, 148-9.
[224] ibid 149.

> '[Online] infrastructure dramatically configures people's options and ends up steering collective action in problematic ways...By enabling only some forms of engagement and positive affectivity, social media "facilitate[es] a web of positive sentiments in which users are constantly prompted to like, enjoy, recommend, and buy as opposed to discuss and critique."'[225]

Essentially, the algorithms employed by online intermediaries allow for users to 'craft their own individual [information] diets.'[226] The consumption of content, and subsequently ideas, on social media sites and other online intermediaries is dominated by selective exposure.[227] The services social networks provide are driven by user data – the way in which users interact with information they are presented 'facilitates audience targeting and personalisation to an unprecedented extent.'[228]

Interaction with Individuals and the Effect on the Idea-Consumer

Given that information access on online intermediaries was originally designed to facilitate entertainment 'accounting for users' preferences and attitudes,'[229] it is unsurprising that idea-consumers are presented with ideas with which they agree. The diffusion of content, and therefore ideas, online is driven by social homogeneity[230] – 'the degree to which preferences of individuals in a society tend to be alike.'[231] The result is the creation of echo chambers.[232]

Even if such intermediaries are not taking an active role in filtering information, their infrastructure can be susceptible to the spread of anti-information. Twitter's infrastructure in particular, due to its building of parasocial/asymmetrical networks of trust, attention feedback loops (where the more attention that is given to an idea by other users the more likely it is to appear on another idea-consumers' view of the

[225] Stefania Milan, 'When Algorithms Shape Collective Action: Social Media and the Dynamics of Cloud Protesting' (2015) 1(2) Social Media + Society 1-10, 8.
[226] Napoli (n 189) 74.
[227] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi and Michele Starnini, 'The echo chamber effect on social media' (2021) 118(9) Proceedings of the National Academy of Sciences of the United States of America 1-8, 5.
[228] Napoli (n 189) 75.
[229] Cinelli et al. (n 222).
[230] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H Eugene Stanley, Walter Quattrociochi, 'The spreading of misinformation online' (2016) 113(3) Proceedings of the National Academy of Sciences of the United States of America 554-559, 558.
[231] William Gehrlein, 'A comparative analysis of measures of social homogeneity' (1987) 21 Quality and Quantity 219-31, 219.
[232] Cinelli et al. (n 222) and Napoli (n 189) 75-6.

information space), and through encouraging declarative statements through limiting the length of messages that can be posted is susceptible to the spread of anti-information.[233]

Although a "by-product" of what makes these intermediaries 'so engaging (and profitable)... [it is] at the expense of critical points of view.'[234] Echo chambers may both "deflect" anti-information that would act as counterspeech to a true idea that is held by the user and also deflect true ideas that would act as counterspeech to anti-information that the idea-consumer believes.[235] The more partisan an idea-consumer's echo chamber, the more segregated they are by the intermediary's information structure, the more likely they are to consume anti-information and resist counterspeech.[236]

Regardless, 'those with an economic and/or political interest in the dissemination of false news are now far better equipped than in the past to deliver their content to those they most desire to reach.'[237]

Chapter 2.3.3: Bots

Bots are automated programs that make decisions without human intervention and are able to adapt to the context in which they operate.[238] There are many types of bot including, crawlers and scrapers,[239] chatbots,[240] spambots and social bots.[241]

Spambots often operate through computers owned by the spammer as well as third party computers which have been hijacked for the purpose of sending spam *en masse*.[242] They are generally used to spread advertisements and malware.[243] They can also be used to perform DDoS attacks.[244] Spambots

[233] See Chamberlain (n 195) for an in-detail explanation.
[234] Rhodes (n 188) 15.
[235] Napoli (n 189) 78.
[236] ibid 79.
[237] ibid 75.
[238] Robert Gorwa and Douglas Guilbeault, 'Unpacking the Social Media Bot: A Typology to Guide Research and Policy' (2020) 12(2) Policy & Internet 225-248, 228.
[239] These bots are indexers and are 'an infrastructural element of search engines and other features of the modern World Wide Web,' which may inflate reports as to how much web traffic is internet usage by a bot, since they are a "bot" despite not directly interacting with users. ibid 229.
[240] These bots 'approximate human speech and interact with humans directly through some sort of interface,' for example, virtual assistants such as Apple's Siri or Amazon's Alexa. ibid 230.
[241] ibid 230-1.
[242] ibid 230.
[243] ibid 231.
[244] ibid 230. DoS (denial-of-service) attacks flood the host of a site or service with requests (asking for information or for a function to be performed) to disrupt the site or service by overloading it, rendering it unavailable for real users. DDoS (distributed denial-of-service) attacks originate from many sources.

can be programmed to search for targets. Spam crawlers search for emails to send spam to and other spambots search for places to post comments and then post automatically.[245] Mass automatic communication would certainly take advantage of the role idea repetition plays in determining an individual's beliefs.

Social bots automatically produce content and interact with humans,[246] mimicking humans,[247] to infiltrate networks of real users,[248] to spread advertisements or malware,[249] and to manipulate public opinion.[250] "Political bots" (social bots for manipulating public opinion)[251] were used during the Brexit referendum campaigns to push hyperpartisan content and 'balkanise readerships,'[252] to affect political discussion in the run up to the 2016 US Presidential Election,[253] to spread anti-information and alt-right narratives in the run up to the 2017 French Presidential Election and likely the 2016 US Presidential Election also.[254] Beyond concerted efforts to make political gains, social bots have also been used for activism,[255] and journalism.[256]

[245] ibid 231.
[246] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini (2016) 59(7) Communications of the ACM 96-104, 94.
[247] ibid. Also, Stefan Stieglitz, Florian Brachten, Björn Ross, and Anna Jung, 'Do Social Bots Dream of Electric Sheep? A Categorisation of Social Media Bot Accounts' (2017) ACIS 2017 Proceedings 89.
[248] Yazan Boshmaf, Ildar Muslukhov, Konstantin Beznosov and Matei Ripeanu, 'The socialbot network: when bots socialize for fame and money' (2011) Proceedings of the 27th Annual Computer Security Applications Conference 93-102, 93 and 99-100.
[249] Gorwa and Guilbeault (n 233) 231.
[250] Samuel C. Woolley and Philip N. Howard, 'Political Communication, Computation Propaganda, and Autonomous Agents' (2016) 10 International Journal of Communication 4882-4890, 4885-7.
[251] ibid 4885.
[252] Marco T. Bastos and Dan Mercea, 'The Brexit Botnet and User-Generated Hyperpartisan News' (2017) 37(1) Social Science Computer Review 38-54, 51-52.
[253] Alessandro Bessi and Emilio Ferrara, 'Social bots distort the 2016 U.S. Presidential election online discussion' (2016) 21(11) First Monday <https://firstmonday.org/ojs/index.php/fm/article/view/7090> accessed 22 April 2023.
[254] Emilio Ferrara, 'Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election' (2017) 22(8) First Monday <https://firstmonday.org/ojs/index.php/fm/article/view/8005> accessed 22 April 2023.
[255] See Heather Ford, Elizabeth Dubois, and Cornelius Puschmann, 'Keeping Ottawa Honest–One Tweet at a Time? Politicians, Journalists, Wikipedians and Their Twitter Bots' (2016) 10 International Journal of Communication 4891-4914.
[256] See Tetyana Lokot and Nicholas Diakopoulous, 'News Bots: Automating news and information dissemination on Twitter' (2016) 4(6) Digital Journalism 682-699.

Bots, with a relatively small number of accounts, are responsible for a large amount of internet traffic that carriers anti-information,[257] particularly because of their role in amplifying such content early on, before it's "viral."[258] Bots also succeed in disseminating anti-information by targeting users vulnerable to manipulation who are likely to repost their content.[259] Bots are more likely to be sharing politically polarised content,[260] not only targeting vulnerable users with the content they share but also engaging with the content the users share, which they modify to cause more discord and "feed the trolls."[261] Bot networks can be taken advantage of by users wishing to share polarising content, just by interacting with the bots.[262] "Low credibility" sources are often heavily supported by bot networks,[263] regardless of the political message of the content.[264]

Since bots are designed to increase discord and to disseminate an equal amount of content on either side of an issue, not engaging with anti-information, by making no attempt to disprove it, could lead to less overall anti-information being shared by bots as no attempt would be made to equalise the discussion.[265]

The volume of bot speech that can be produced is a large part of the issue with them.[266] A bot can emulate a human user whilst operating much quicker than a human could,[267]and bots are able to operate consistently continuing to outpace humans even when human discussion peaks.[268] Those with the

[257] Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini & Filippo Menczer, 'The spread of low-credibility content by social bots' (2018) 9 Nature Communications 5 <https://www.nature.com/articles/s41467-018-06930-7> accessed 22 April 2023.
[258] ibid.
[259] ibid. Also, Marina Azzimonti and Marcos Fernandes, 'Social media networks, fake news, and polarization' (2023) 76 European Journal of Political Economy 23 <https://www.sciencedirect.com/science/article/pii/S0176268022000623> accessed 22 April 2023.
[260] David A. Broniatowski, Amelia M. Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C. Quinn and Mark Dredze, 'Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate' (2018) 108(10) American Journal of Public Health 1378-1384, 1383.
[261] ibid.
[262] ibid 1382.
[263] Shao et al. (n 252) 5.
[264] Ibid. Also, Broniatowski (n 255) 1383.
[265] Broniatowski (n 255) 1383.
[266] Erin Griffith, 'Pro-Gun Russian Bots Flood Twitter After Parkland Shooting' (WIRED, 15 February 2018) <https://www.wired.com/story/pro-gun-russian-bots-flood-twitter-after-parkland-shooting/> accessed 16 April 2023.
Max de Haldevang, 'Russian trolls and bots are flooding Twitter with Ford-Kavanaugh disinformation' (Quartz, 2 October 2018) <https://qz.com/1409102/russian-trolls-and-bots-are-flooding-twitter-with-ford-kavanaugh-disinformation> accessed 16 April 2023.
[267] Tim Hwang, Ian Pearce, and Max Nanis, 'Socialbots: Voices from the Fronts' 19(2) Interactions 38-45, 40-41.
[268] Bessi and Ferrara (n 253).

capabilities to run large bot networks can kick up 'massive clouds of claims, accusations, misinformation, and controversies, they can overwhelm the capacity of the public and traditional media to respond to any of them; thus causing a type of paralysis.'[269]

Bots, like targeted advertisements, may be used to overwhelm the idea-consumer. Their ability to appear and act as though they are humans means that they can have the same effects a human speaker might. Many bots may be operated by an individual human speaker, and therefore they may multiply the effect that human speaker could have on their own. Bot networks can cause ideas to appear truthful simply by virtue of the number of "people" that believe it.

Chapter 2.4: Implications for Idea-consumption

Part 1 of this chapter outlined the argument from truth and marketplace of ideas as truth-seeking theories and questioned their applicability to facts, as opposed to debatable ideas. Part 2 of this chapter then demonstrated that the idea-consumer, the agent that supposedly performs the truth-seeking function that the marketplace of ideas (and the argument from truth) describes is not a rational actor and is often unable to discern factual truths and falsity. The way the idea-consumer selects information to base their beliefs upon, the way the idea-consumer obtains beliefs, and the way the idea-consumer evaluates ideas all differ from the rational consumer. The idea-consumer, by not always favouring the truth as a rational idea-consumer would, is not guaranteed to perform its necessary function in the marketplace of ideas of discerning truth. Part 3 of this chapter demonstrated how different technologies contribute to the dissemination of anti-information. Alongside the psychological barriers the idea-consumer faces, a "more speech" solution appears absurd. Different solutions to the anti-information issue are discussed in the next chapter.

[269] On flooding tactics, applicable to bots due to the volume of speech they are able to create – Zeynep Tufekci, 'Twitter and Tear Gas: The Power and Fragility of Networked Protest' (Yale University Press 2017) 274.

Chapter 3: Forming an Effective Anti-Information Strategy

This chapter considers various approaches to the anti-information issue in light of the discussion in the previous chapter. This determines the strengths and limitations of each and leads to the conclusion that a strategy that prevents anti-information entering the marketplace of ideas in the first place is the preferred approach to the issue. Where prevention fails, or where it cannot be appropriately applied because of overriding speech interests, the appropriate strategy should be chosen from the remaining – removal/access control, information correction, and more speech.

This chapter will conclude that prevention is the preferred approach. There is no need for an anti-information strategy if there is no anti-information in the marketplace of ideas. However, given that it is likely anti-information, in some form, will always exist in the marketplace of ideas, it is necessary to be able to determine what approach is appropriate in a given situation to abate its harms. Doing so will entail closer examination of the strengths of each approach.

This previous chapter demonstrated that reliance solely upon a "more speech" approach, the solution favoured by the truth-seeking theories, is insufficient to address the anti-information issue. This section will connect the psychological disadvantages of the idea-consumer to the technological advantages anti-information and its disseminators hold. In doing so, it will argue that the preferred strategy for addressing anti-information should prevent anti-information from entering the marketplace of ideas, then the appropriate remedy of removal/limitation of access, information correction and "more speech" should be applied where prevention fails. It will start by considering the different solutions in turn, starting with "more speech."

Chapter 3.1: More Speech

Certain technologies and characteristics of the (online) marketplace of ideas exacerbate the psychological barriers the idea-consumer faces.

Targeted advertisements take advantage of a number of psychological weaknesses of the idea-consumer. Given that the idea-consumer obtains beliefs/consumes ideas in a passive manner, and can do so through repetition of ideas, targeted advertisements can disseminate anti-information very effectively. Given that the idea-consumer need not engage with that idea to come to believe it and that the more they see it the more likely they are to believe it, targeted advertisements gain effectiveness through taking advantage of those weaknesses. An advertisement that delivers a message designed to be effective against a certain individual is already disproportionately effective compared to other advertisements. However, a targeted advertisement can take advantage of confirmation bias, by targeting those who fall within a demographic likely to already believe in a false idea.

Similarly, bots take particular advantage of the illusory truth effect, where repeated encounters with an idea causes the individual to believe it, and the false consensus effect, where an individual believes a false idea they hold is in the majority. By flooding social media feeds with anti-information, repeatedly exposing the idea-consumer to the same anti-information and giving the appearance that many people support an idea, bots/bot networks in their ability to post mass communications take advantage of these weaknesses.

Likewise, by exposing the idea-consumer to many more ideas with which they agree than disagree, algorithms and echo chambers reinforce some evaluative biases. Notably, it supports the false consensus effect and confirmation bias. This is connected to the issue of the economic incentive for intermediaries allowing for anti-information to be disseminated, and even encouraged. More information allows for platforms to tailor the content they show in a way that pleases the user. There is no opportunity for the idea-consumer to come to believe counterspeech because they do not encounter it, and when the algorithm fails and they do, they are unlikely to come to believe it because of biases in the selection of ideas upon which the idea-consumer bases their beliefs.

"More speech" is hardly a solution to the problem when there is plenty of adequate ideas available but none being believed. There is not a scarcity of ideas but an inability to process that which are encountered and filter out the bad.[1] "More speech" solutions often take the form of transparency measures and opposing viewpoint discussion.

It is often assumed that transparency-improving measures are helpful for combating anti-information,[2] despite a lack of sufficient evidence to support its effects.[3] 'Predictions about the inevitable increase of understanding and democratization of the world through transparency are often narrowly conceived and in many cases wrong.'[4] Transparency should be seen as a means to combating anti-information, as opposed to an end itself.[5]

The assumed effectiveness of transparency often rests on another assumption that individuals seek out and correctly interpret information shared by transparency-improving measures;[6] the onus is on idea-consumers to make use of information shared due to transparency norms.[7] There is an assumption that idea-consumers are actively engaged with the information made available through transparency-improving measures and that they have the capacity to interpret that information correctly. The assumption that access to information "speaks for itself" is weak.[8]

In some instances, the average idea-consumer will be engaged and will have capacity. For example, for nutritional labels – all consumers that are concerned with the nutritional content of their food will

[1] Derek Bambauer, 'Shopping Badly: Cognitive Biases, Communications, and the Fallacy of the Marketplace of Ideas' (2006) 77(3) University of Colorado Law Review 649-710, 696-7.
[2] For example, a survey by Marília Gehrke found 80.6% of participants agreed that transparency is helpful for countering anti-information (although this survey was limited to Brazilian journalists) – Marília Gehrke, 'Transparency as a key element of data journalism: perceptions of Brazilian professionals' (Computation + Journalism Symposium, 2020) available at <https://cpb-us-w2.wpmucdn.com/sites.northeastern.edu/dist/d/53/files/2020/02/CJ_2020_paper_8.pdf> accessed 2 November 2022.
[3] Tim Wood and Melissa Aronczyk, 'Publicity and Transparency' (2020) 64(11) American Behavioral Scientist 1531-1544, 1537.
[4] Lars Thøger Christensen and George Cheney, 'Peering into Transparency: Challenging Ideals, Proxies, and Organizational Practices' (2015) 25(1) Communication Theory 70-90, 84, referring to Kristin Lord, *The Perils and Promise of Global Transparency: Why the Information Revolution May Not Lead to Security, Democracy or Peace* (State University of New York Press 2006).
[5] ibid 71, and Wood and Aronczyk (n 3).
[6] Sabina Schnell, 'Transparency in a "Post-Fact" World' (2022) 5(3) Perspectives on Public Management and Governance 222-231, 222.
[7] Wood and Aroncyzk (n 3) 1537, Christensen and Cheney (n 4) 71 and Schnell (n 6) 222.
[8] Christensen and Cheney (n 4) 71.

engage with nutritional labels. Not all consumers will have the capacity to interpret the ingredients and nutritional information. Nutritional labels do not illuminate much about the food production process.[9] This could be similarly applied to industry standards requiring transparency regarding CO2 emissions. Whilst they provide some insight, transparency 'may simultaneously obscure more complex questions related to the issue of how reductions are accomplished, for example through trading of emissions.'[10]As such, making acts of transparency "legible" becomes the work of 'journalists, marketers, social media users, communications professionals, government agencies, and others.'[11]

Transparency of raw information is not particularly helpful as it does not address the issue of the idea-consumer's lack of rationality.[12] Additionally, raw information can be manipulated and distorted by "strategic actors" to take advantage of the biases of idea-consumers,[13] that are 'based on pre-existing beliefs, emotions and identities.'[14]

Opposing viewpoint discussion, where a voice is given to opposing views from the source, is an approach that would be compliant with the argument from truth and the marketplace of ideas. It is a literal "more speech" approach, where alternative ideas are encountered alongside the original idea. It could be considered preventative in the sense that any anti-information that may be disseminated would be opposed and therefore, presumably, gain no adherents. Yet there is evidence to the contrary, beyond that which has been discussed already.

Information sources that have opposing viewpoint discussion, as opposed to discussing only one side of a particular issue, have higher perceived levels of credibility but only if the source is ideologically supportive.[15] Opposing viewpoint discussion resulted in reduced use of liberal media outlets by conservative viewers.[16] A conclusion drawn from these findings was that conservative media consumers,

[9] Christensen and Cheney (n 4) 78.
[10] ibid.
[11] Wood and Aroncyzk (n 3) 1538.
[12] Schnell (n 6) 224.
[13] ibid 225.
[14] ibid 224-5.
[15] Jay Hmielowski, Sarah Staggs, Myiah Hutchens, and Michael Beam, 'Talking Politics: The Relationship Between Supportive and Opposing Discussion With Partisan Media Credibility and Use' (2022) 49(2) Communications Research 221-244, 238.
[16] ibid.

through years of conservative media consumption, 'may have hit a point of homeostasis relative to their media credibility perceptions.'[17] That is to say, they are unlikely to find opposing viewpoints credible unless they come from a conservative media source. Contrastingly, the 'more trusting of media' liberals are susceptible to changes in their media attitudes.[18]

Opposing viewpoint discussion, a literal more speech approach which is applied from the first instance an idea is shared, is only conditionally effective for helping idea-consumers discover the "truth." Where speech with opposing viewpoint discussion enters the marketplace of ideas it is unlikely to be more effective, given the role of psychological biases and technology. Given that a "more speech" approach alone cannot be relied upon, the remaining solutions are information correction, removal and access control, and prevention.

Additionally, a "more speech" approach allows for certain individuals to have a disproportionate effect on the marketplace of ideas.

Habermas considered the purpose of public discourse to be 'the creation of "a common will, communicatively shaped and discursively clarified."'[19] For the common will to have any legitimacy, discourse must be '"immunised against repression" and "all force" must be excluded, "except the force of the better argument."'[20]

Force has been interpreted as including wealth. Gouldner echoing Habermas says, 'the "rationality of 'public' discourse... depends on the prior possibility of separating speakers from their normal powers and privileges in the larger society, especially in the class system, and on successfully defining these powers and privileges as irrelevant to the quality of their discourse."'[21]

[17] ibid 239.
[18] ibid.
[19] Jürgen Habermas (trs), *The Theory of Communicative Action* (1987) 81-2, in Robert Post, 'The Constitutional Concept of Public Discourse: Outrageous Opinion, Democratic Deliberation, and Hustler Magazine v. Falwell' (1990) 103(3) Harvard Law Review 601-86, 640.
[20] Habermas (n 19) 25-6.
[21] Alvin Gouldner, *The Dialectic of Ideology and Technology: The Origins, Grammar and Future of Ideology* (1976) 98, in Post (n 19) 641.

Wealth, particularly the ability to purchase targeted advertisements, has given anti-information the chance to 'disrupt the marketplace of ideas in entirely new and troubling ways.'[22] Idea-consumers using social media intermediaries are "overwhelmed" and have a "disrupted sense of reality,"[23] particularly with political speech where they may be exposed to 'a high volume of misinformation and conspiratorial content than professionally produced news.'[24]

Influence also acts as a force on the marketplace of ideas. Many idea-consumers are on the weaker end of a power imbalance, compared to certain speakers, such as the relationship between advertisers and economic consumers. There is no area where this is quite so concerning as in political ideas. "Ideas" may be regulated to protect against harm where there is a power imbalance, such as through consumer protection laws, but political ideas are heavily protected, and harder to regulate even where there is harm.

The role of influence in affecting the idea-consumers evaluation of an idea,[25] and the role of wealth in allowing a speaker access to means that are otherwise unavailable means that wealthy and influential individuals exercise a force disproportionate to the "truth" of their idea on the marketplace of ideas.

A "more speech" solution, whilst favoured by the truth-seeking theories, faces a number of difficulties in countering anti-information. This section demonstrated how different technologies and biases may interact in a way that prevents a "more speech" approach being effective. Additionally, it discussed transparency and opposing viewpoint discussion as "more speech" solutions. It was concluded that whilst transparency is assumed to be effective at countering anti-information, it often is not because the idea-consumer must be capable of understanding the information which they encounter, and lots of "raw info" is illegible to the average person, whilst being capable of being manipulated by strategic actors. Further, opposing viewpoint discussion fails to be an effective "more speech" solution as it needs information to come from an ideology/attitude-consistent source. Finally, this section highlighted the

[22] Daniela Manzi, 'Managing the Misinformation Marketplace: The First Amendment and the Fight Against Fake News' (2019) 87(6) Fordham Law Review 2623-2651, 2628.
[23] ibid.
[24] Phillip Napoli, 'What If More Speech Is No Longer the Solution? First Amendment Theory Meets Fake News and the Filter Bubble' (2018) 70 Federal Communications Law Journal 55-104, 73.
[25] See Chapter 2.2.4.

role of wealth, by giving individuals a greater capability to speak in the marketplace of ideas, particularly through new technologies, and influence, due to the trust people place in influential people as "idea-rating agents," in preventing "more speech" solutions being effective as the speech of wealthy and influential speakers often has disproportionate effects on the marketplace of ideas than it should by virtue of its "truth."

Chapter 3.2: Information Correction

Information correction would include attaching disclaimers to posts that spread anti-information on social media, ensuring corrections are made by newspapers for anti-information that is published, and robustly fact-checking public officials. Additionally, it would include "prebunking," where idea-consumers are "inoculated" by messages that pre-empt anti-information and refute it.[26] An information correction approach is consistent with the truth-seeking justifications for freedom of expression;[27] adding more information only helps individuals determine the truth. Beyond doing nothing, it is the least restrictive option available.[28] However, an information correction strategy faces similar challenges to a "more speech" approach.

As with a more speech approach, information correction may worsen confirmation bias. It has been found that information correction is more likely to backfire where the information concerns contentious topics, where factual claims are ambiguous and where the correction strategy is not robust enough.[29] An insufficiently robust information correction strategy may be one which designates posts as fake but does little to actually correct the information – a mistake that Facebook made.[30]

[26] See for example, Stephan Lewandowsky & Sander van der Linden, 'Counter Misinformation and Fake News Through Inoculation and Prebunking' (2021) 32(2) European Review of Social Psychology 348-384, 357.
[27] Rebecca Helm and Hitoshi Nasu, 'Regulatory Responses to 'Fake News' and Freedom of Expression: Normative and Empirical Evaluation' (2021) 21(2) Human Rights Law Review 302-28, 315.
[28] ibid.
[29] Helm and Nasu (n 27) 317.
[30] Andras Koltay, 'Constitutional protection of lies?' (2020) 25(3) Communications Law 131-149, 143.

Even when an individual believes a correction, anti-information continues to influence their perception of an event.[31] The ability of beliefs to persevere despite attitude-inconsistent information can be particularly strong,[32] therefore, information correction may be insufficient to combat anti-information where particularly strong biases are present, for example, confirmation bias.

Lack of economic incentive for platforms to reduce the presence of anti-information could particularly be an issue with regards to an information correction strategy. Industry self-regulation remains unlikely so an information correction strategy would have to be imposed on platforms. The structure of an information correction strategy could potentially be problematic if online platforms are left to be the arbiters of truth. Issues of censorship are not so easily abated by reliance upon an "independent fact checker" either, as false fact checkers have caused issues for information correction strategies.[33]

Additionally, beyond concerns of censorship and ineffectiveness due to psychological barriers of the idea-consumer, information correction faces a practical issue in the volume of speech it would have to address. Lack of gatekeeping over speech, and the usage of bots means that large volumes of anti-information may flood the marketplace of ideas, and inhibit discussion and its truth-seeking purpose.

This practical issue however may not be such a barrier to the effectiveness of an information correction strategy where repetitive speech, that takes advantage of the illusory truth effect, is disseminated. Repeating information correction would be no more difficult than repeating anti-information, and could be done automatically, the difficulty would be in detecting discrete pieces of anti-information to be refuted.[34]

[31] Michael Cacciatore, 'Misinformation and public opinion of science and health: Approaches, findings and future directions' (2021) 118(15) Proceedings of the National Academy of Sciences of the United States of America 1-8, 4.
[32] See Moti Nissani and Donna Marie Hoefler-Nissani, 'Experimental Studies of Belief Dependence of Observations and of Resistance to Conceptual Change' (1992) 9(2) Cognition and Instruction 97-111, 103-6.
[33] See Andrew Moshirnia, 'Who Will Check the Checkers? False Factcheckers and Memetic Misinformation' (2020) 4 Utah Law Review 5.
[34] Nicollas de Oliveira, Pedro Pisa, Martin Andreoni Lopez, Dianne Scherly de Medeiros and Diogo Mattos, 'Identifying Fake News on Social Networks Based on Natural Language Processing: Trends and Challenges' (2021) 12(1) Information 38, 2.

An information correction approach would require specific focus on how correction interacts with biases,[35] how to ensure against a possibility of censorship and how to deal with the technological difficulties that are raised in identifying anti-information amongst mass communication. Whilst information correction offers an alternative idea which the idea-consumer can believe instead of anti-information, there is no guarantee they will.

Chapter 3.3: Removal and Access Control

Preventing content that contains anti-information from being accessed, entirely or partially, is the most restrictive solution discussed so far and it is the first that is incompatible with traditional approaches to the truth-seeking theories. Removing content or limiting access to it is the "enforced silence" the described by Justice Brandeis in *Whitney*.[36]

Limiting access to anti-information may take the form of warning the user that it is anti-information before they access it and having them confirm they understand that they are to encounter something identified as anti-information.[37] An approach which limits the access to content identified as anti-information should avoid "shadowbanning," where users have their content restricted without being informed so that it is not publicly available or not available without searching for it, due to the restriction it places upon future expression,[38] and the secrecy of the measure. As with information correction, this highlights the need for effective detection of anti-information.

A removal/access approach addresses many technological effects as well as the selection and unconscious biases directly. In particular, changing the likelihood users will encounter anti-information works to prevent users entering an echo chamber and reduces the effects of anti-information as a whole.

Additionally, the effects of the selection and unconscious biases in causing an idea-consumer to believe anti-information are limited by this approach because the idea-consumer would be presented with more

[35] Helm and Nasu (n 27) 306.
[36] *Whitney v California* [1927] 274 US 357, 377 (United States of America). See Chapter 2.1.2.
[37] Assuming issues with the process of identifying information as anti-information are first resolved, see Chapter 2.1.3.
[38] See for example, *Lingens v Austria* App. No.9815/82, [44].

truthful ideas and the information they encounter online would not be limited by algorithms solely to information that appeases their biases.

However, this approach evidently runs a similar risk to information correction in that if applied incorrectly it could act as censorship. Caution as to what information has its access limited or is removed from an online platform would be advised, and transparency would be necessary.

A particular risk removal of anti-information faces is incidentally legitimising the information in the eyes of the idea-consumer. The "Streisand effect" is 'a social phenomenon whereby the removal of content can actually draw increased attention to it.'[39] Further, removal of false stories can serve as "proof" of conspiracy or suppression of truth to those committed to it.[40] Simple removal of popular stories is therefore likely to be ineffective as the information would have already spread much further than the original source and because removal could potentially worsen its effects.[41] This effect could be worsened if removal occurs repeatedly to the same information, as may happen if trolls or bots propagate anti-information widely.

Conversely, early detection and removal of anti-information before bot networks amplify it and make it go viral,[42] could be extremely effective. Though since identification of bots relies upon spotting "bot characteristics,"[43] anti-information that originates from the mistake of a real person,[44] may not be able to be prevented earlier, as the content of the message may appear human. If bots make effective use of anti-information created or disseminated by users, early detection of bot-originating anti-information would prove pointless.

---

[39] Helm and Nasu (n 27) 321. See also Sue Curry Jansen and Brian Martin, 'The Streisand Effect and Censorship Backfire' (2016) 9 International Journal of Communication 656-71.
[40] Helm and Nasu (n 27) 321. See also, for the effect occurring with news designated as "fake" by Facebook, Koltay (n 30) 143.
[41] Helm and Nasu (n 27) 321.
[42] See Chapter 2.3.3.
[43] See Berta García-Orosa, Pablo Gamallo, Patricia Martín-Rodilla and Rodrigo Martínez-Castaño, 'Hybrid Intelligence Strategies for Identifying, Classifying and Analyzing Political Bots' (2021) 10(10) Social Sciences <https://www.mdpi.com/2076-0760/10/10/357> accessed 13 June 2023. See also, Rachit Shukla, Adwitiya Sinha and Ankit Chaudhary, 'TweezBot: An AI-Driven Online Media Bot Identification Algorithm for Twitter Social Networks' (2022) 11(5) Electronics <https://www.mdpi.com/2079-9292/11/5/743> accessed 13 June 2023.
[44] For example, RaeAnn Christensen, 'Austin man says sorry for posting misleading anti-Trump protester Tweet' (FOX 4 News, 14 November 2016) <https://www.fox4news.com/news/austin-man-says-sorry-for-posting-misleading-anti-trump-protester-tweet> accessed 3 June 2023.

Regardless, presumably one effect of the removal/limitation of anti-information and promotion of truthful information would be any false consensus that may have been created by bots, algorithms and echo chambers could be prevented. If anti-information is removed from a platform effectively, all that (or the majority that) remains would be the truthful information upon which ideas may be based. The idea-consumer would be forced to confront their naïve realism,[45] and to truly "test" their thinking in the marketplace of ideas.[46]

Chapter 3.4: Prevention

Anti-information holds many advantages over the idea-consumer due to psychological weaknesses in the selection of information and the obtention and evaluation of ideas. Additionally, anti-information thrives in the online marketplace of ideas where its dissemination is economically incentivised and there is no gatekeeping of speaking. More speech, information correction and removing/access limitation are all flawed solutions to the issue.

The only solution which is completely effective is preventing the dissemination of anti-information in the first place,[47] particularly because many of the effects of anti-information are felt from the idea-consumer's first encounter with it.[48] This could take the form of criminal sanction.[49] However, not every instance of anti-information will be prevented and therefore, the correct approach(es) between more speech, information correction and removal/access limitation must be employed where prevention fails.

A "more speech" approach evidently does not interfere with speech. "More speech" however lacks the ability to adequately address the problem, given that psychological biases and effects and new technologies prevent "more speech" solutions like transparency and opposing viewpoint discussion from working effectively. Additionally, it does not address the disproportionate force wealthy and influential people may have over the marketplace of ideas. The idea-consumer alone is unable to

[45] See Chapter 2.2.4.
[46] *Abrams v US* [1919] 250 US 616, 630 (United States of America).
[47] Helm and Nasu (n 27) 326.
[48] See Chapter 2.2
[49] Helm and Nasu (n 27) 322-5.

perform the truth-seeking function that is envisioned for them, and as such, speech interfering approaches to the anti-information issue like information correction, removal/access limitation and prevention must be employed. Though, in some circumstances where there is a sufficiently strong speech interest, or where restriction upon speech is disproportionate to the aim of protecting against anti-information a "more speech" solution may be appropriate.

Information correction and removal/access limitation may be appropriate in some circumstances where preventing a type/form of speech entering the marketplace of ideas is disproportionate to the aim of protecting against anti-information. Though in employing either strategy, to ensure they are effective, their usage should be informed by the latest research on how such solutions interact with psychological biases and effects, and research on how to best employ them in communication systems (such as social media sites).

Chapter 4: Directly Addressing Anti-Information

As will be discussed, the Honest Ads Bill and the Online Safety Bill have been touted as solutions to the anti-information issue. Therefore, this Chapter considers the proposals of both and their appropriateness for addressing the harms anti-information presents.

The first section discusses the "Honest Ads" approach. The Honest Ads Bill requires disclaimers on political advertising detailing who paid for the advert. It concludes that whilst the "Honest Ads" approach may be appropriate given First Amendment jurisprudence, it is not appropriate for addressing the anti-information issue. It finds that the Honest Ads Bill and similar approaches, as "more speech" approaches are of little utility – particularly because by not addressing the content of speech itself, disseminating anti-information in contravention of such a provision may be "worth it." This section does not consider compatibility of the Honest Ads Bill or a measure like it with the ECHR because a provision within the Elections Act 2022 employs an "Honest Ads" approach, although it has not yet entered into force.

The second section discusses the Online Safety Bill. The Online Safety Bill has a large scope however, this section discusses the Terms of Service duties that apply to online platforms, and the false communication offence that it creates. It concludes that with regards to anti-information the Online Safety Bill does little to move beyond industry self-regulation, particularly due to the heavy reliance on the TOS and transparency duties and that although the false communication offence represents a step towards a preventative approach, its inapplicability to certain forms of anti-information mean that it is insufficient to protect against all the harms anti-information presents.

Chapter 4.1: Honest Ads

This section consists of four parts. The first part outlines the current state of election advertising regulation in the US. This part demonstrates the need for an effective regulator by highlighting the inadequacies of the FEC. The second part discusses the Honest Ads Bill and finds that its effectiveness is likely dependent upon how it is enforced. The third part discusses the possibility for "Honest Ads" in the UK through the proposed extension of imprinting requirements to cover online political advertisements. It concludes that the extension of imprinting requirements, whilst holding advantages over the Honest Ads Bill, suffers from the same weaknesses in that it is a "more speech" solution. The fourth part considers whether an "Honest Ads" approach is an appropriate solution to the anti-information issue. It finds that whilst the Honest Ads Bill is appropriate in the US, measures should go further in the UK.

Chapter 4.1.1: Present US Regulation and the FEC

This part briefly discusses the FEC, the current state of election advertising regulation in the US and the FEC's role in regulating election advertising.

Federal Election Commission

The FEC is the campaign finance regulator in the US. It was created through amendments to the Federal Election Campaign Act of 1971.[1] It administers the Federal Election Campaign Act and has 'exclusive jurisdiction with respect to the civil enforcement of [it],'[2] therefore is able to initiate civil actions to enforce the provisions of the Act.[3] Its powers include obtaining evidence under oath, requiring the attendance of witnesses by subpoena, and conduct investigations,[4] after a compliant is filed by a person believing the Act has been violated or if the Commission itself believes a violation has been committed.[5]

[1] Federal Election Campaign Act Amendments of 1974, Pub L 93-433; 52 US Code 30106 (US).
[2] 52 US Code 30106(b).
[3] 52 US Code 30107(a)(6).
[4] 52 US Code 30107(a).
[5] 52 US Code 30109(a).

Present US Election Advertising Regulation

The Honest Ads Bill seeks to extend reporting requirements for electioneering communications online.

An electioneering communication is 'any broadcast, cable or satellite communication' which refers to a particular candidate, and that which, depending on the type of election, is within 30-60 days of that election, and is targeted at the relevant electorate.[6] The 1971 Act requires any legal or natural person who "produces and airs electioneering communications in excess of $10,000 in a calendar year," and any legal or natural person who has an independent expenditure of at least $1,000 within a certain number of days of an election to file a report to the FEC.[7]

Additionally, the Honest Ads Bill seeks to extend disclosure requirements online, to include public communications[8] within disclosure requirements and to heighten the threshold of disclosure.

Currently, the Federal Election Campaign Act requires the speaker "clearly state" who paid for the communication,[9] and whilst it has specific requirements for print,[10] radio,[11] and television,[12] these requirements also do not extend online as it applies to public communication and electioneering communications. The definition of each does not include online communications.

Application of Existing Rules

Winichakul notes that existing rules could apply to political communication on the internet, but only fail to due to narrow interpretation of those rules by the FEC.[13] Its interpretation of rules outlining what groups are required to report to the FEC results in many persons and communications escaping regulation.[14] This sentiment is echoed by Jacobs, who notes that the gaps in transparency laws is what

[6] 52 US Code 30104(f)(3)(A).
[7] 52 US Code 30104(f)(1).
[8] Paid political communications made to a general audience which excludes online communication – 52 US Code 30101(22).
[9] 52 US Code 30120(a).
[10] 52 US Code 30120(c).
[11] 52 US Code 30120(d)(1)(a).
[12] 52 US Code 30120(d)(1)(b).
[13] Pichaya Winichakul, 'The Missing Structural Debate: Reforming Disclosure of Online Political Communications' (2018) 93(5) New York University Law Review 1387, 1394.
[14] ibid.

limits their effectiveness,[15] and similarly refers to the FEC's interpretation of its current campaign disclosure rules.[16]

The ineffective enforcement of campaign finance rules is worsened by a lack of enforcement action by the FEC,[17] as well as a failure to update rules[18] or account for technological changes.[19] 'The absence of a threat of enforcement and punishment has changed the law-abiding norms of regulated entities – new political groups now brazenly disobey the law.'[20]

The structural issues of the FEC are seemingly to blame. The Chair and Vice-Chair of the Commission is changed yearly, with only one year in each role every six years allowed for each of the six commission members.[21] Further, since the Chair and Vice-Chair may not be from the same party,[22] 'the Chair rotates annually not only from person-to-person but also from party-to-party.'[23] Such a system prevents leadership from actively improving and influencing the Commission's agenda.[24] Voting on draft enforcement action and policy opinions also causes major set-backs with procedure being reinitiated upon any objecting vote.[25] This can be exploited by those wishing to cause ambiguity in the regulations through delay in their enforcement by simply causing the Commission to have to make a decision on a particular issue.[26]

These issues highlight the need of, not only fixing substantive issues where regulation fails to cover internet communications, but of ensuring proper enforcement through an effective regulator. Although the structure of the Federal Electoral Commission is much different to the UK's Electoral Commission, it would be wrong to assume this problem is distinctly an American one. Especially given, the recent

[15] Leslie Gielow Jacobs, 'Freedom of Speech and Regulation of Fake News' (2022) 70(S1) The American Journal of Comparative Law i278-i311, i296.
[16] ibid i296 cf Abby Wood and Ann Ravel, 'Fool Me Once: Regulating "Fake News" and Other Online Advertising' (2018) 91(6) Southern California Law Review 1223-1278, 1249-50.
[17] ibid 1398.
[18] ibid 1397 'little rulemaking activity regarding Internet communications since 2006.'
[19] ibid 1403.
[20] ibid 1407.
[21] 52 US Code 30106(a)(2)(D)(5).
[22] ibid.
[23] Winichakul (n 13) 1401.
[24] ibid.
[25] ibid 1402.
[26] ibid 1410.

changes to the Electoral Commission,[27] which have been criticised for threatening the Electoral Commission's independence and accountability in politics.[28] An ineffective regulator may be exploited to allow for anti-information to continue to thrive and is therefore an issue worth anticipating.

Chapter 4.1.2: The Honest Ads Bill

At present the reporting requirements apply only to television, radio and print media, however, the Honest Ads Bill proposes to extend the requirements to online election advertising, and the Honest Ads Bill proposes to strengthen the disclaimer requirements for all forms of advertising.

The Honest Ads Bill has been introduced in the US Senate twice, in October 2017 and May 2019.[29] A companion bill was introduced in the House of Representatives also in May 2019.[30] The first part of this section will discuss the present US regulation on reporting and disclaimer requirements for election

[27] A number of changes have been made by the Elections Act 2022. Firstly, the Act allows for the Secretary of State to direct the Electoral Commission's work through a 'strategy and policy statement' (s16, cf s4A, Political Parties, Elections and Referendums Act 2000), of which the Electoral Commission have a 'duty to have regard to' (s16 cf s4B PPERA2000) performance of which the Speaker may examine (s17 cf s13ZA PPERA2000). Additionally, it allows for any Minister to replace the Secretary of State for Levelling Up, Housing and Communities in the Speaker's Committee on the Electoral Commission under s2, PPERA2000, the body responsible for criticism the Electoral Commission (s18 cf s2(2A) PPERA2000). Finally, the Act also prevents the Electoral Commission from instituting criminal proceedings in England and Wales or Northern Ireland (s19, cf Schedule 2, PPERA2000). The sections of the Elections Act 2022 which made these changes came into force on 19 August 2022, see The Elections Act 2022 (Commencement No.1 and Saving Provision) Regulations 2022, SI 2022/908.

[28] The Electoral Commission, 'A strategy and policy statement for the Electoral Commission' (5 July 2021) <https://www.electoralcommission.org.uk/who-we-are-and-what-we-do/our-views-and-research/elections-act/a-strategy-and-policy-statement-electoral-commission> accessed 30 September 2022.
Nick Cohen, 'The Tories call it electoral reform. Looks more like a bid to rig the system.' (Guardian, 18 December 2021) <https://www.theguardian.com/commentisfree/2021/dec/18/the-tories-call-it-electoral-reform-looks-more-like-a-bid-to-rig-the-system> accessed 30 September 2022.
Anita Bhadani, 'Democracy fears following "authoritarian" grab of Electoral Commission' (The National, 28 April 2022) <https://www.thenational.scot/news/20101860.democracy-fears-following-authoritarian-grab-electoral-commission/> accessed 30 September 2022.
Alistair Clark, 'Elections Bill: a modest proposal to improve the Speaker's Committee on the Electoral Commission' <https://consoc.org.uk/elections-bill-a-modest-proposal-to-improve-the-speakers-committee-on-the-electoral-commission/> accessed 30 September 2022.
The Electoral Commission, 'The Electoral Commission's ability to bring prosecutions' (5 July 2021) <https://www.electoralcommission.org.uk/who-we-are-and-what-we-do/our-views-and-research/elections-act/electoral-commissions-ability-bring-prosecutions> accessed 30 September 2022.

[29] Congress, S.1989 – Honest Ads Act, <https://www.congress.gov/bill/115th-congress/senate-bill/1989/actions> accessed 15 January 2022.
Congress, S.1356 - A bill to enhance transparency and accountability for online political advertisements by requiring those who purchase and publish such ads to disclose information about the advertisements to the public, and for other purposes, <https://www.congress.gov/bill/116th-congress/senate-bill/1356/actions> accessed 15 January 2022.

[30] Congress, H.R.2592 – Honest Ads Act, <https://www.congress.gov/bill/116th-congress/house-bill/2592/actions> accessed 15 January 2022.

advertising. Additionally, it will explain the role of the Federal Election Commission (the regulator) and the issues it faces. The second part of this section will then cover the Honest Ads Bill's proposals, criticisms it faces and, through the ..., its likely effectiveness.

The Bill targets social media sites but also search engines like Google.[31] The Bill is a transparency-improving measure which 'is fundamental to [American] democracy... [allows the electorate] to make informed political choices and... is essential to enforce other campaign finance laws, including the prohibition on campaign spending by foreign nationals.'[32]

Proposal

The Honest Ads Bill would amend the 1971 Act to include within its reporting online communication by extending the definition of "electioneering communications" to 'any communication which is placed or promoted for a fee on an online platform.'[33] Similarly, it expands the definition of public communication to online communications by including within its definition 'paid internet, or paid digital communication,'[34] which alongside the extended definition of electioneering communications, extends disclosure requirements online.

Additionally, Section 7 of the Honest Ads Bill would heighten the disclosure threshold by replacing 'clearly state' will 'state in a clear and conspicuous manner.'[35] Also, it makes specific provision for the disclosure requirements of online communications, similarly to the current provisions for print, radio and television.[36]

[31] The text is the same across all three introductions of the Bill.
Text of 2017 Senate Bill, <https://www.congress.gov/bill/115th-congress/senate-bill/1989/text> accessed 15 January 2022.
Text of 2019 Senate Bill, <https://www.congress.gov/bill/116th-congress/senate-bill/1356/text> accessed 15 January 2022.
Text of 2019 House of Representatives Bill, <https://www.congress.gov/bill/116th-congress/house-bill/2592/text> accessed 15 January 2022.
Henceforth, the Honest Ads Bill.
See Sec.8(a) re the subsection to be added to the 1971 Act. The definition provided for online platform includes search engines and the definition for 'qualified political advertisement' includes 'search engine marketing.'
[32] ibid Sec.4.
[33] ibid Sec.6(a)(1).
[34] ibid Sec.5.
[35] ibid Sec.6.
[36] ibid.

The guidance in the Bill becomes a minimum standard that disclosure statements would have to meet. For example, it requires that disclosures in video communications must be at least 4 seconds long and appear in writing and in an audible format,[37] to prevent quickly spoken barely audible disclosures as well as small-print illegible disclosures.

Finally, the Bill adds record keeping requirements for "online platforms" into the 1971 Act. "Online platforms" are defined in Section 8(3) of the Bill as 'any public-facing website, web application, or digital application (including a social network, ad network, or search engine) which – sells qualified political advertisements; and has 50,000,000 or more unique monthly United States visitors or users for a majority of months during the preceding 12 months [of an election].'[38] The record is to contain a copy of the advertisement, a description of the audience targeted by the advertisement, and information about the rate charged for the advertisement and requests made regarding it.[39]

Criticisms

One structural factor limiting the Bill's scope is the requirement of online platforms to have 50 million unique monthly users to be subject to the Bill. The US population is approximately 330 million, 302 million of which are internet users. A requirement of approximately a sixth of all internet users in the US to use a particular website each month for it to qualify as an "online platform" for the purposes of the Honest Ads Act excludes all but the most popular websites – 'barely 20 websites in the US qualify.'[40]

Although such websites make up a small percentage of those used by American idea-consumers, they may expose more idea-consumers to more political advertisements than the unaffected platforms. However, this may still be a significant limiting factor. Despite catching sites like Facebook and Twitter,

[37] ibid Sec.7(b)(1).
[38] Honest Ads Bill (n 31).
[39] ibid Sec.8(2).
[40] Doug Zanger, 'Industry Opinion: Is the Honest Ads Act a viable solution for digital political advertising?' (The Drum, 24 October 2017) <https://www.thedrum.com/news/2017/10/24/industry-opinion-the-honest-ads-act-viable-solution-digital-political-advertising> accessed 27 August 2022.

it ‘leaves thousands of other platforms viewed by hundreds of millions of Americans open to foreign propaganda.’[41]

Another limitation of the Bill’s ability to address the anti-information issue is the limited effect of record keeping. As with other transparency-improving mechanisms raw information needs to be made “legibile” for the idea-consumer, and this process can be manipulated by bad faith actors. Although the record keeping requirement would aid the FEC in enforcing related laws,[42] it is difficult to imagine an unengaged user devoting time to examining a database to learn more about who paid for a post. ‘Not to mention then taking the time to decide whether they should believe it or not based on the list.’[43] Therefore, its effect is likely only incidental. It will come down to ‘the FEC’s appetite for enforcement’ and its use of the record keeping “tool.”[44]

Whilst transparency through record keeping, ‘won’t, by itself, solve the problem of election interference’ [45] or of anti-information generally, optimistically, it could mean ‘that those minds susceptible to positioning tactics, rhetoric, or outright exploitation won't be as easily moved once they understand the true origins and intent of the ads.’[46] Although a speaker still does not need to ‘“spin” the story through honesty’[47] as the content of their speech remains beyond the scope of Bill, disclosure and transparent record keeping at least encourages expression to be made from a point of good faith and limits the effect of wealth on the marketplace of ideas. At the very least, more true information will be available to the idea-consumer.[48]

Finally, the Bill’s scope is significantly limited by its inapplicability to the majority of anti-information. The Bill is “ill-suited” to addressing misleading expressions since as it only targets advertising, it

[41] Zanger (n 40).
[42] Honest Ads Bill (n 31) Sec.4.
[43] Zanger (n 40).
[44] Ellen Goodman and Lyndsey Wajert, ‘The Honest Ads Act Won’t End Social Media Disinformation, But It’s A Start’ (unpublished) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3064451> accessed 26 August 2022.
[45] Zanger (n 40).
[46] ibid.
[47] ibid.
[48] According to the American marketplace of ideas, if not our own.

'overlooks the volume of internet-enabled "cheap speech."'[49] Ironically, despite the main concern regarding anti-information in the US being foreign interference in campaigns,[50] the Bill fails to address 'Russian propaganda... posted on free social media platforms.'[51]

The Honest Ads Bill would also be seen as only a "start"[52] by those that argue for industry self-regulation on top of legislative action, including Beyersdorf who says that it is difficult to draft laws that 'reach elusive trolls and bots and to cover content outside the traditional scope of campaign finance regulation without infringing on the First Amendment.'[53] Soupcoff agrees that legislation will never be broad enough to capture all the disseminators of anti-information but also 'too broad to avoid chilling the speech of normal Americans wanting to express a political opinion.'[54] To describe the Honest Ads Bill as chilling appears unconvincing given the US Supreme Court's finds disclosure requirements are 'a less restrictive alternative to more comprehensive regulations of speech.'[55]

Beyersdorf believes the Honest Ads Bill is 'key... especially if the FEC fails to overcome its partisan divide' and because the voluntarily adopted policies of social media companies 'remain imperfect.'[56] As such the act would provide a "level playing field" for social media companies and would give the US the ability to enforce the requirements of the act to ensure a high standard of effective disclosure.[57]

Likely Effectiveness

Ultimately, the effectiveness of the Honest Ads Bill would depend upon its enforcement. Nonetheless, should it be effectively enforced, it is a significant limitation on wealth as an influential force on the marketplace of ideas. Those that employ substantial finances to disrupt the online marketplace of ideas

[49] Tawanna Lee, 'Combating Fake News with "Reasonable Standards"' (2021) 43(1) Hastings Communications and Entertainment Law Journal 81-107, 102. See also Wood and Ravel (n 16) 1249-50.
[50] Jennifer M Grygiel, 'Algorithmic propaganda: how Facebook meddles with democracy' (2020) 25(1) Communications Law 23-30, 23.
[51] Lee Goodman, '"Honest" political ads: Watch out, Drudge, you're next' (The Hill, 4 September 2019) <https://thehill.com/opinion/cybersecurity/459896-honest-political-ads-watch-out-drudge-youre-next/> accessed 27 August 2022.
[52] Goodman and Wajert (n 16).
[53] Brian Beyersdorf, 'Regulating the "Most Accessible Marketplace of Ideas in History": Disclosure Requirements in Online Political Advertisements After the 2016 Election'(2019) 107(3) California Law Review 1061-1100, 1098-9.
[54] Marni Soupcoff, 'Honest Ads in the Agora' (2017) 40 Regulation 80, 80.
[55] *Citizens United v Federal Election Commission* (2010) 558 US 310, 369 (United States of America).
[56] Beyersdorf (n 53) 1091.
[57] ibid 1094.

would be prevented from doing so anonymously. Whilst the Federal Election Commission would be able to enforce the rules within the Bill itself, the added transparency could be a useful tool for journalists that wish to question the motive behind an advertisement, which may in turn lower its utility. Additionally, its lack of applicability to "cheap speech" is a significant limiting factor, especially given the role bots and people with influence can play in the dissemination of anti-information.

This approach is preventative in the sense that it seeks to discourage dishonest advertisement however, since the Bill does not address the content of the speech it regulates, there is nothing that actually prevents anti-information entering the marketplace of ideas. Instead, it is merely a more speech approach which relies upon a currently ineffective regulator and journalists.

Nonetheless, it is not completely undesirable. A preventative approach is generally preferred but where that fails other strategies are necessary. Improving transparency and following a more speech approach by making campaign finance rules on disclosure stricter, covering all paid political communications online, would help to eliminate, or at least discourage for fear of political repercussions, the use of "dark money."[58]

Chapter 4.1.3: Honest Ads in the UK

The Elections Act 2022 makes provision for the extension of "imprinting" requirements to online electoral communications, as recommended by the Law Commission and Scottish Law Commission,[59] though the relevant provisions have not yet entered into force, and require regulation made by the Secretary of State to do so.[60]

At present, Section 110 the Representation of the People Act 1983 requires for printed material the inclusion of the name and address of the '(a) printer of the document; (b) promoter of the material and

[58] Jacobs (n 15) i296. Lachlan Markay and Andrew Desiderio, 'How Gridlock, Social Media Giants and the Clintons Made the Internet Ripe for Russian Meddling' (The Daily Beast, 20 October 2017) <https://www.thedailybeast.com/how-gridlock-social-media-titans-and-the-clintons-turned-the-internet-into-the-wild-west-of-american-politics> accessed 4 May 2023.
[59] Law Commission and Scottish Law Commission, 'Electoral Law: A joint final report' (Law Com No 389, Scot Law Com No 256, 2020) paras 11.70-11.72.
[60] Elections Act 2022, s67.

(c) any person on behalf of whom the material is being published.'[61] This is mirrored in the Elections Act, bar information as to the printer.[62]

The extension of imprinting requirements applies to electronic material, that is not a telephone call or SMS message,[63] which 'consists of or includes – (a) text or moving or still images, or (b) speech or music.'[64] This is further limited by the purpose served by the electronic material, which differs for paid-for electronic and other electronic material.[65]

Paid-for electronic material[66] requires an imprint where it is influencing the public or any section of the public to give or withhold support from a registered party,[67] any registered parties that do or do not advocate a certain policy,[68] any candidates that do or do not hold certain opinions or advocate a certain policy,[69] or any registered parties or candidates that can otherwise be categorised,[70] any material that is influencing the public or any section of the public to give or withhold support from any candidate,[71] or office-holder,[72] or office-holders that do or do not advocate a certain policy or hold a certain opinion.[73] As well as any material that is influencing the public or any section of the public to give or withhold support from the holding of a referendum,[74] or the outcome of such a referendum.[75]

Other electronic material[76] must wholly or mainly relate to a referendum,[77] or must promote or procure electoral success for a registered party and candidates and parties holding certain opinions,[78] as with

[61] Representation of the People Act 1983, s110(3).
[62] Elections Act 2022, s41(3).
[63] ibid s39(3).
[64] ibid s39(2).
[65] ibid ss42-5.
[66] ibid s42.
[67] ibid s43(2)(a).
[68] ibid s43(2)(b).
[69] ibid s43(2)(c).
[70] ibid s43(2)(b)-(c).
[71] ibid s43(4).
[72] ibid s43(6).
[73] ibid s43(7).
[74] ibid s43(9)(a).
[75] ibid s43(9)(b).
[76] ibid s44.
[77] ibid s44(2)(b).
[78] ibid s45(2)(a)-(c).

paid-for electronic material, promote or procure the election of a particular candidate,[79] or the success or failure of a recall petition.[80]

Breaching the imprinting requirement constitutes an offence for the promoter of the material and any person on behalf of whom the material is published,[81] for which on summary conviction the guilty person is liable to a fine.[82] A court may order that the material is removed or access to it is disabled,[83] failure to comply with the order is also an offence,[84] with the same punishment.[85]

Structural Strengths

Like the Honest Ads Bill, it is limited to specific purposes. However, there are convoluted and asymmetrical standard being applied depending on the type of vote, and the material being imprinted which cannot be justified under the goal of increasing transparency in digital advertising.[86]

Unlike the Honest Ads Bill however, it does apply to "cheap speech" by including material other than paid-for electronic material. This approach is preferable to that taken by the Honest Ads Bill as it addresses the disproportionate force influence may have on the marketplace of ideas, whilst also addressing the role of wealth in purchasing misleading advertisements. Additionally, there is no reliance upon a regulator and instead the provisions are to be enforced through prosecution.[87]

Likely Effectiveness

Principally, in order to be effective, these provisions must enter into force. It seems likely that they will as one relevant provision to online imprinting requirements which requires guidance to be produced on the operation of that part of the act has entered into force.[88] Its likely effectiveness is similar to that of

[79] ibid s45(5).
[80] ibid s45(7).
[81] ibid s48(1). Or in the case of material the purpose of which is to influence the public to give or withhold support from a particular candidate, a person would instead by guilty of an illegal practice – s48(8) cf Schedule 11 and the Representation of the People Act 1983.
[82] ibid s48(2).
[83] ibid s49(2).
[84] ibid s49(4).
[85] ibid s49(5).
[86] Law Commission and Scottish Law Commission (n 59) 11.71.
[87] Elections Act 2022, ss57-9.
[88] The Elections Act 2022 (Commencement No. 8) Regulations 2023, SI 2023/552, cf Elections Act 2022, s54.

the Honest Ads Bill. Importantly, this proposal has a broader scope than the Honest Ads Bill, covering cheap speech as well as paid-for advertising. As with other transparency measures, it is likely only effective if the transparent information can be made "legible" to the ordinary person. Additionally, it does not address the issue of limited time. Should an advertisement be placed in contravention of this provision just before an election, it may not be remedied in time. The punishment being a fine might be a sum wealthy actors that wish to influence elections are willing to pay.

Chapter 4.1.4: Appropriateness of an Honest Ads Approach

The Honest Ads Bill and the First Amendment

Whether a measure affecting free speech is content-based or content-neutral is an important element of First Amendment jurisprudence. Determining content-neutrality was once considered to be a question of 'whether the government has adopted a regulation of speech because of disagreement with the message it conveys.'[89] Justice Marshall, in *Police Department of the City of Chicago v Mosley*, did not frame the question as requiring (dis)agreement with the message conveyed by the speech – 'the First Amendment means that government has no power to restrict expression because of its message, its ideas, its subject matter, or its content.'[90]

Thus, First Amendment jurisprudence finds measures that discriminate against a particular idea or viewpoint to be 'an egregious form of content discrimination.'[91] This understanding was reaffirmed recently. In 2015, in *Reed v Town of Gilbert, Arizona*, where the "commonsense meaning" of content-based was said to require a court to consider whether a measure 'draws distinctions based on the message a speaker conveys.'[92] If a measure is content-based, it 'applies to particular speech because of the topic discussed or the idea or message expressed'[93] and 'is subject to strict scrutiny regardless of

[89] *Ward v Rock Against Racism* [1989] 491 US 781, 791 (United States of America).
[90] [1972] 408 US 92, 95 (United States of America).
[91] Per Justice Kennedy, *Rosenberger v Rector and Visitors of University of Virginia* [1995] 515 US 819,829 (United States of America).
[92] [2015] 576 US 155, 156 (United States of America).
[93] ibid 163.

the government's benign motive, content-neutral justification, or lack of "animus towards the ideas contained" in the regulated speech.'[94]

Principally, the Honest Ads Bill does not address the content of disseminated ideas. It mostly relies upon reporting and record keeping. The Bill itself 'does nothing to make ads "honest."'[95] It does place restrictions on the manner of expression in that it must be accompanied by a disclosure.

Having this limitation is understandable from the legislator's perspective, given that a measure taken to address the content of speech would be unlikely to survive a First Amendment challenge. '[T]he First Amendment means that government has no power to restrict expression because of its message, its ideas, its subject matter, or its content.'[96] Although, some content-based legislation may be constitutional, given that it survives strict scrutiny.[97] Strict scrutiny is the most stringent form of judicial review in the US. It is rare for a speech restriction to survive strict scrutiny,[98] which requires a compelling Government interest addressed through the least restrictive means which is narrowly tailored to that interest.[99] Therefore, to affect the content of political advertisements, as opposed to how those advertisements are made, would likely trigger a successful challenge.

Limiting the requirements of the Bill to mandating disclosure increases its likelihood of surviving a First Amendment challenge given the judgment in *Citizens United v FEC*, which reaffirmed the US Supreme Court's position that disclosure requirements are 'a less restrictive alternative to more comprehensive regulations of speech.'[100] Such requirements provide the information necessary for

[94] ibid 156 and 165.
[95] Bradley Smith, 'Misnamed "Honest Ads Act" would restrict free speech' (USA Today, 12 June 2019) <https://eu.usatoday.com/story/opinion/2019/06/12/election-interference-honest-ads-act-threatens-free-speech-editorials-debates/1438271001/> accessed 27 August 2022.
[96] *Police Department of Chicago* (n 90) 95 (United States of America), per Justice Marshall.
[97] *Reed v Town of Gilbert* (n 92) 182-3.
[98] *Williams-Yulee v Florida Bar* [2015] 575 US 433, 444.
[99] *Sable Communications v Federal Communications Commission* [1989] 492 US 115, 126.
[100] *Citizens United v Federal Election Commission* [2010] 558 US 310, 369.

voters to make informed decisions.[101] Nunziato agrees,[102] and King states that 'the bill's language reflects the drafters' cognizance of Supreme Court precedent.'[103]

Extension of Imprinting Requirements and Convention Compliance

Principally, political advertising falls within the scope of Art.10's protection,[104] for which there is no European consensus on how it should be regulated, invoking a wide margin of appreciation.[105]

There is no content-neutrality requirement under Art.10. Restrictions of protected expression based upon content are permissible so long as they are prescribed by law, in pursuit of a legitimate aim and necessary in a democratic society. For example, convictions for gratuitously offensive speech.[106] Though greater protection would be afforded to speech affected by the provisions in the Elections Act, due to the heightened level of protection that political speech receives.[107]

There is nothing about the Elections Act that would suggest it fails to meet the requirements of "prescribed by law" that being, it is accessible and its consequences are sufficiently foreseeable.[108]

The State could pursue the legitimate aim of the protection of the rights of others, namely the right to free elections, as guaranteed by Article 3 of Protocol 1 of the Convention,[109] specifically 'the need to protect the electoral process as part of the democratic order.'[110] For which it would receive a wide margin of appreciation since it is 'better placed to assess the difficulties faced in establishing and safeguarding the democratic order.'[111]

[101] This was the position held by the majority and some dissenting judges in *McConnell v Federal Election Commission* (2003) 540 US 93, which was referenced in coming to the decision in *Citizens United.*
[102] Dawn Carla Nunziato, 'The Marketplace of Ideas Online' (2019) 94 Notre Dame Law Review 1519, 1556. For example, *Buckley v Valeo* [1976] 424 US 1, 66-68.
[103] John King, 'Microtargeted Political Ads: An Intractable Problem'(2022) 102(3) Boston University Law Review 1129-67, 1157.
[104] See for example, *Animal Defenders International v UK*, App. No.48876/08.
[105] ibid [123].
[106] *ES v Austria*, App. No.38450/12, [43].
[107] *Wingrove v UK*, App. No.17419/90, [58].
[108] *The Sunday Times v UK (No.1)*, App. No.6538/74, [49].
[109] To which the UK is a signatory.
[110] *Animal Defenders International v UK* (n 103) [111].
[111] *Ždanoka v Latvia* App. No.58278/00, [134].

Although the risk to the right to free elections caused by "dishonest" ads is evidently greater immediately preceding and during electoral periods,[112] the risk is not 'confined to such periods since the democratic process is a continuing one to be nurtured at all times by a free and pluralist debate.'[113] The Honest Ads Bill limits its effects to the period preceding an election through the definition of electioneering communication. Similarly, French Law 2018-1202 *on the fight against the manipulation of information* which allows judges to order 'all proportionate and necessary measures to stop [the] dissemination [of anti-information]'[114] operates in the three months preceding the first day of the month of a general election.[115]

Whilst an act that limits its operation to the period preceding an election is more likely compliant, an act which does not like the Elections Act is not unjustifiable, particularly as the difficulty faced in removing belief in anti-information would suggest there is a need for the democratic process to receive ongoing protection.[116]

The necessity of the interference requires that there is a pressing social need and the restriction must be proportionate to the legitimate aim. The pressing social need could be demonstrated by reference to the harms anti-information presents,[117] the surge of spending on online advertisements,[118] and the relevant experiences with regards to targeted advertisements recently in the UK (for example, the one billion adverts the Leave Campaign sent days before the Brexit Referendum). The restriction – the imposition of transparency disclosures on political advertising – is a minimal one which is likely proportionate to the legitimate aim. To be considered proportionate, there must not be less restrictive means of achieving the aim pursued.[119] Transparency requirements are the least restrictive measure and are a solution

[112] *Bowman v UK* App. No.24839/94, [43].
[113] *Animal Defenders* (n 104) [111].
[114] LOI n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information (1), article 1, translation by author, available at <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000037847559> accessed 24 April 2023.
[115] ibid.
[116] See Chapter 2.2 and Rachael Craufurd Smith, 'Fake news, French Law and democratic legitimacy: lessons for the United Kingdom?' (2019) 11(1) Journal of Media Law 52-81, 59.
[117] Joint Committee on the Draft Online Safety Bill, *Draft Online Safety Bill* (2021-22, HL 129, HC 609) para 35.
[118] The Electoral Commission, 'Know who is paying for online political ads' <https://www.electoralcommission.org.uk/i-am-a/voter/online-campaigning/know-who-paying-online-political-ads> accessed 23 April 2023.
[119] *Glor v Switzerland* App.No.13444/04, [94].

favoured by the Committee of Ministers of the Council of Europe 'so as to enable [the public] to form an opinion on the value to be given to information, ideas and opinions disseminated.'[120] Referring in particular to the possible distortion of debate that could occur due to groups with large financial power buying political advertisements online, and the risk to "free and pluralist debate,"[121] would demonstrate the necessity of the interference.[122]

Extension of imprinting requirements by the Elections Act is very likely Convention compliant.

Honest Ads as a Means to Protect Against Anti-Information

The "Honest Ads" approach is an appropriate measure for protecting against anti-information under the First Amendment. It strikes a reasonable balance between the speech interests at stake and the need to protect against anti-information, in the context of the First Amendment.

However, it is a weak measure. Critically, it does not directly address the anti-information issue as it does not address the content of speech itself. Although the proposal in the UK covers more than the Honest Ads Bill, it is still limited in the harms it can be applied to, particularly with regards to cheap speech. Therefore, both conceptions suffer from limited applicability to the issue.

Whilst it is respectful of speech interests, as a solution to the anti-information issue, it has little utility. Beyond issues with transparency-improving approaches, as have already been discussed, its weak sanctions mean that disseminating anti-information may still be "worth it." Particularly given the strength of anti-information in the electoral context, where time to reveal the truth through discussion is limited and where the idea-consumer rates many ideas vicariously through the opinions of influential speakers that they trust.

A more appropriate measure would be stricter or would address the content of speech itself.

[120] Council of Europe, Committee of Ministers, Recommendation CM/Rec(94)13 on measures to promote media transparency (22/11/1994), 'General provisions on media transparency.' Also, states should encourage initiatives that 'improve the transparency of the process of online distribution of media content, including automated process' – Council of Europe, Committee of Ministers, Recommendation CM/Rec(2018)1 on media pluralism and transparency of media ownership, [2.5].
[121] *Animal Defenders* (n 104) [112].
[122] ibid [112] and [122].

Chapter 4.2: The Online Safety Bill

Protection from online harms is mostly at present a self-regulated responsibility of providers of online services. The Online Safety Bill 'will end the era of self-regulation'[123] through a "platform-based"[124] system which places duties of care on online service providers. It has been suggested that the imposition of duties is a continuation of the "self-regulatory approach" which has failed so far.[125] Whereas others have welcomed the scheme and either have concerns about its effect on free speech,[126] or that it does not address particular harms that may be encountered online, such as the effects of algorithms,[127] disinformation,[128] and AI.[129]

The *Online Harms* White Paper, published in April 2019, found that 'the existing "patchwork of regulation and voluntary initiatives" had not gone far or fast enough to keep UK users safe. It therefore proposed a single regulatory framework to tackle a range of online harms.'[130] The proposals of this White Paper were developed into the Online Safety Bill which remains before Parliament. The Online Safety Bill's framework splits duties between different service providers, categorised primarily by whether it is a "user-to-user service" or a "search service."

Disinformation has repeatedly been stated to be a harm for which the Online Safety Bill will provide protection.[131] However, the Bill does not explicitly state that misinformation or disinformation will be considered "harmful," and the scope of the duties to address harmful content have been narrowed. Nonetheless, other aspects of the Bill are relevant to the issue.

[123] HL Deb 18 May 2021, vol 812, col 517.
[124] ibid col 502.
[125] ibid cols 473-4.
[126] ibid cols 490, 497-8, 504, 507.
[127] ibid col 544. Also, HC Deb 11 May 2022, vol 714, col 215.
[128] HC Deb 11 May 2022, vol 714, col 215.
[129] ibid.
[130] Briefing Paper, Regulating Online Harms, Number 8743, 28 May 2021.
[131] See for example, Department for Digital, Culture, Media and Sport, *Online Harms White Paper: Full Government Response to the Consultation* (CP 354, December 2020) [34].

This section will detail which services are captured by the proposed regulatory scheme, the way the approach to harmful content has changed, the false communication offence the Bill introduces, and the appropriateness of the Bill in protecting against anti-information.

The Online Safety Bill was amended the day before thesis submission, therefore, the following is correct as of 21 June 2023.

Chapter 4.2.1: Ofcom and the Regulated Services

This section will briefly cover Ofcom, the regulator, and the services that the Bill seeks to regulate.

Ofcom

Ofcom, the Office of Communications, was established by Section 1 of the Office of Communications Act 2002. However, its functions and duties are mostly contained within the Communications Act 2003. Ofcom's principal duty is 'to further the interests of citizens in relation to communications matters.'[132] Ofcom also has a duty to promote media literacy,[133] which will be discussed further in Chapter 4.2.

Under the Online Safety Bill, Ofcom has a number of duties and powers relevant to enforcement of the Bill and subsequently, to the anti-information issue. Section 119 of the Bill details the duties in the Bill that Ofcom can enforce through notices of contravention.[134] When a provisional notice of confirmation is confirmed,[135] Ofcom may require that a person take particular steps,[136] pay a penalty,[137] or both.[138] The Bill contains penalties for failure to comply with Ofcom's decisions.[139]

Additionally, the Bill grants Ofcom the power to issue information notices,[140] the power to require interviews,[141] and powers of entry, inspection and audit,[142] to gather information about compliance

[132] 'and; (b) to further the interests of consumers in relevant markets, where appropriate by promoting competition.' Communications Act 2003, s3(1)(a)-(b).
[133] Communications Act 2003, s11.
[134] Online Safety Bill HL Bill (2022-23) 87(Rev) s118.
[135] ibid s120.
[136] ibid s120(5)(a) cf s121.
[137] ibid s120(5)(b) cf s125.
[138] ibid s120(5)(c).
[139] ibid s126-28.
[140] ibid s91.
[141] ibid s96.
[142] ibid s97 cf Schedule 12.

whilst imposing a duty to co-operate with any investigations Ofcom opens.[143] Similarly, there are offences to comply with information notices and investigations.[144]

Finally, Section 139 of the Bill requires that Ofcom establishes and maintains an advisory committee on disinformation and misinformation.[145] Which must publish a report within 18 months of establishment and then periodically after that.[146]

Regulated Services

The services with which the Bill is concerned are "internet services."[147] Meaning simply that which are 'made available by means of the internet.'[148] The primary distinction between services that the Bill seeks to regulate is "user-to-user services" compared to "search services." The Bill describes a "user-to-user" service as an internet service where user-generated content,[149] may be shared and encountered by another user,[150] regardless of what proportion of the content on that service is user-generated, so long as user-generated content may be shared.[151] Whereas a "search service" is an internet service which includes a search engine,[152] particularly that 'service or functionality which enables a person to search some websites or databases'[153] so long as it enables a person to search more than just one website or database.[154]

[143] ibid s95(1).
[144] ibid s98-102.
[145] ibid s139(1).
[146] ibid s139(5).
[147] ibid s200.
[148] ibid s200(1). Including services made available by mobile data s200(2)(b) and s200(3) cf Communications Act 2003, s32(2).
[149] The Bill defines "user-generated content" as "content" that is generated on the service by a user, or that which is uploaded to/shared on the service by a user which may be encounter by other users through the service (s49(3)). This includes content which is generated, uploaded or shared by a bot (s49(4)).
"Content" receives a broad description and includes 'anything communicated by means of an internet service, whether publicly or privately... data of any description' (s207(1)).
[150] Online Safety Bill (n 135) s2(1).
[151] ibid s2(2).
[152] ibid s2(4).
[153] ibid s204(3).
[154] ibid s204(2)(b) cf plurality of 'websites' and 'databases' in s204(3).

There are 6 main duties imposed on the regulated services (any user-to-user or search service, that 'has links with the United Kingdom,'[155] except those which are exempt).[156] There are some differences between the extent of duties for user-to-user services and for search services, for example: user-to-user services must notify Ofcom where a risk assessment identifies the presence of content harmful to children which is not covered by the then definition of harmful under the Bill.[157]

A further distinction is drawn regarding the category under which a service falls. A duty is imposed upon Ofcom to establish a register of the categorisation which sorts the services into Category 1, Category 2A and Category 2B as soon as reasonably practicable, as well as a duty to maintain the register.[158] Category 1 is for user-to-user services only,[159] whilst Category 2A is for search services and user-to-user services which include a search engine,[160] and Category 2B is for user-to-user services.[161] Schedule 11 says that the category a service will be registered as depends upon its number of users and its functionalities which the Secretary of State will specify as well as other factors that the Secretary of State considers relevant.[162] Subsequently, the Secretary of State effectively decides to which services duties apply and the extent to which they apply.

Although we do not know what specific criteria will be used to categorise a service, the number of users being one suggests that the duties limited to Category 1 services, to be discussed in this section, are intended for the most popular sites. It is safe to assume the likes of Facebook, YouTube and Twitter will be categorised as Category 1 services, but where the cut-off point is and whether less popular services like Discord, Reddit and Tumblr which still have millions of UK users,[163] would be categorised as Category 1 services is unknown.

---

[155] ibid s3(2)(a). A service has links with the UK 'if – (a) the service has a significant number of United Kingdom users, or (b) United Kingdom users form one of the target markets for the service (or the only target market)' (s3(5)).
[156] ibid s3(2)(b). Exemptions are provided for in Schedule 1 of the Bill.
[157] ibid s10(5) cf. s24. See also, 'non-designated content' s54(6).
[158] ibid s87(1)-(3).
[159] ibid s87(1).
[160] ibid s87(2). A user-to-user service which includes a search engine is referred to as a "combined service" in the Bill (s3(7)).
[161] Online Safety Bill (n 135) s87(3).
[162] ibid Schedule 11, Paragraph 1(1)-(3).
[163] 10% of messenger users in the UK use Discord – Statista, 'Discord brand awareness, usage, popularity, loyalty, and buzz among messenger users in the UK in 2022.'

Chapter 4.2.2: Changing Approach to Harmful Content

Early criticism of the White Paper's proposals suggested that it blurred the line between law and morality by moving between "illegal" and "unacceptable" content.[164] Earlier drafts of the Bill created a "legal but harmful" category of speech which raised concern over whether it 'will create a situation in which people are prevented from saying things that are legal but prohibited.'[165] The current draft has removed the harmful content duties that existed in relation to adults,[166] in favour of transparency measures,[167] and user empowerment.[168]

In older drafts, risk assessment duties existed for children and for adults which included assessing the risk of 'priority content.'[169] The Secretary of State was empowered to specify what content fell under 'priority content' under certain conditions.[170] The risks identified in the risk assessments, priority content *inter alia*, were to be mitigated under the safety duties.[171] This content is the "legal but harmful" speech which raised free speech concerns.[172] Having removed the risk assessment duties and corresponding safety duties for adults, the Bill now seeks to protect against harms through user empowerment and transparency with respect to the regulated content.

A term regularly used throughout the Bill is "regulated user-generated content." The Bill defines "user-generated content" as "content" that is generated on the service by a user, or that which is uploaded to/shared on the service by a user which may be encountered by other users through the service.[173] This includes content which is generated, uploaded or shared by a bot.[174] "Regulated user-generated content"

<https://www.statista.com/forecasts/1328633/discord-messengers-brand-profile-in-the-uk> accessed 13 June 2023.

[164] Paul Wragg, 'Tackling online harms: what good is regulation?' (2019) 2 Communications Law 49-51, 49-50

[165] HC Deb 11 May 2022, vol 714, col 189.

[166] See for example, Online Safety HC Bill (2022-23) [209], henceforth "Earlier OSB."

[167] PBC (Bill 209) 13 December 2022, cols 24-25, 86-92, 94-96.

[168] ibid col 50.

[169] Earlier OSB (n 166) s10(6)(b)(ii) and s12(5)(b).

[170] ibid s54-56, 004 draft s53-55.

[171] ibid s11(2)-(10) and s13(2)-(7).

[172] For example, Carla Lockhart, MP for Upper Bann, HC Deb 19 April 2022, vol 712, col 117.

[173] Online Safety Bill (n 135) s49(3).

[174] ibid s49(4).

means user-generated content except for a few excluded types, including (a) emails, (b) SMS, and (c) MMS, *inter alia*.[175]

Terms of Service and Transparency duties

Category 1 services under the Bill have a duty not to act against users except in accordance with their terms of service (TOS).[176] Their duties relating to TOS are relevant to their transparency duties.

There are further duties relating to Category 1 services' TOS, including ensuring that provisions for taking down or restricting access to regulated user-generated content, and for suspending or banning a user from the service, 'are – (i) clear and accessible, and (ii) written in sufficient detail to enable users to be reasonably certain whether the provider would be justified in taking the specified action in a particular case...'[177] Section 65(3) requires services providers to comply with their own TOS.

Additionally, they must operate the service 'using systems and processes that allow users and affected persons' to report both 'relevant content' and persons they believe should be suspended or banned based upon the TOS.[178] "Relevant content" being that which the TOS dictates action will be taken against.[179] Similarly, there must be a complaints procedure allowing for such reporting,[180] and where a person believes the service is not complying with their duties to act against certain content or certain users,[181] and to allow for reporting.[182] The complaints procedure must also allow for users whose content has been affected,[183] or whose account has been suspended or banned to complain.[184]

Yearly, Ofcom must give service providers notice requiring them to produce a transparency report about the services they provide.[185] Schedule 8 of the Bill provides further detail as to the matters about which information in the transparency report may be required by the notice. User-to-user services may be

[175] ibid s49(2).
[176] ibid s64.
[177] ibid s65(4)(a) and (b) they must apply the TOS consistently.
[178] ibid s65(5).
[179] ibid s67(5).
[180] ibid s65(8)(a).
[181] ibid s65(8)(b).
[182] ibid s65(8)(b).
[183] ibid s65(8)(c).
[184] ibid s65(8)(d).
[185] ibid s68(1)-(2).

required to provide information about the "incidence" and "dissemination" of "relevant content,"[186] and the number of users that encountered "relevant content,"[187] the application of the TOS,[188] the systems and processes for users to report and for the provider to deal with "relevant content,"[189] and the "functionalities" to help users manage risks relating to "relevant content."[190] Ofcom itself must also produce transparency reports that summarises those they receive from service providers.[191]

Whilst not explicitly addressing misinformation (policies) in their TOS,[192] Facebook and Twitter both have misinformation policies.[193] Should large social media sites incorporate anti-information into their TOS, anti-information would become "relevant content" and the duties within this section would apply to it.

The TOS and transparency duties are enforceable by Ofcom,[194] and the advisory committee on disinformation and misinformation is to provide advice to Ofcom on providing notice to produce a transparency report and what information to require.[195] Given the extensive powers of investigation, penalisation, and requirement to take steps that Ofcom are to be given under the Bill, the TOS and transparency approach could be effective, but only if service providers choose to include anti-information within their TOS.

[186] ibid Schedule 8 Part 1 Paragraphs 1-2.
[187] ibid Schedule 8 Part 1 Paragraph 3.
[188] ibid Schedule 8 Part 1 Paragraph 4.
[189] ibid Schedule 8 Part 1 Paragraph 5-6.
[190] ibid Schedule 8 Part 1 Paragraph 7.
[191] ibid s145.
[192] Facebook, 'Terms of Service' <https://m.facebook.com/terms/> accessed 1 May 2023.
Twitter, *Twitter User Agreement* <https://cdn.cms-twdigitalassets.com/content/dam/legal-twitter/site-assets/privacy-policy-new/Privacy-Policy-Terms-of-Service_EN.pdf> accessed 1 May 2023.
[193] Meta, 'Misinformation' <https://transparency.fb.com/policies/community-standards/misinformation> accessed 1 May 2023.
Twitter, 'How we address misinformation on Twitter' <https://help.twitter.com/en/resources/addressing-misleading-info> accessed 1 May 2023.
[194] Online Safety Bill (n 135) s119(2).
[195] ibid s139(4)(b).

Effectiveness of Terms of Service Measures

The impetus is on the service providers to implement TOS that address anti-information. Depending on the strength of action they take, this approach could represent a preventative, or a removal/access limitation approach.

With significantly strong measures, disseminators of anti-information may be dissuaded from using the platform altogether. Though it is likely there would still be some dissemination, as beyond blocking the speaker from accessing the service, there is little more the service provider can do, and given the effects encountering a piece of anti-information only once can have,[196] there may still be a significant incentive to disseminate anti-information.

Therefore, a TOS approach would likely manifest as a removal/access control approach to the anti-information issue. Such an approach particularly aids with protecting against the role of influential speakers and technology in the marketplace of ideas. Speakers with large platforms may be less likely to disseminate anti-information, lest they risk losing that platform (if the TOS were sufficiently strict). Further, given that means of mass communication, like bots and targeted advertisements, are often only available to those with resources, those means may be made less effective if access to that speech is impeded by effective enforcement of TOS.

For this online safety issue, the industry remains self-regulatory. However, if platforms are encouraged to adopt strict measures to address anti-information, and Ofcom is able to act as an effective regulator, this approach could be preventative. For this approach to be successful, both platforms in adopting measures within their TOS to address anti-information and Ofcom in monitoring that activity should refer to psychological and communication research to avoid effects like the Streisand effect,[197] that may occur.

[196] See Chapter 2.2.
[197] See Chapter 3.3.

User empowerment duties

The user empowerment duties apply only in relation to Category 1 services.[198] The main duty is to include 'features which adult users may use or apply if they wish to increase their control over [the specified] content...' [199] which includes content that promotes suicide, self-harm and behaviour associated with eating disorders,[200] as well as abuse targeted at race, religion, sex, sexual orientation, disability, or gender reassignment,[201] and content that incites hatred against people due to those characteristics.[202] Those features should trigger systems or processes which reduce the likelihood of the user encountering such content.

Given that there is no possibility for expansion of the duty to further content on the wording of the Bill, there is no possibility this could cover anti-information, unlike under the definition of "harmful" in previous drafts of the Bill. There may be some incidental effect if abuse is based upon anti-information. For example, xenophobic bullying in relation to COVID-19.[203]

Regardless, the effectiveness of user empowerment measures depends ultimately on the number of users that make use of the features. Approximately 80% of users adjust their privacy settings at some point

[198] Online Safety Bill (n 135) s12(1).
[199] ibid s12(2).
[200] ibid s12(10).
[201] ibid s12(11). Notably, the Bill uses 'gender reassignment' rather than "gender identity" excluding transgender people that have not undergone 'a process (or part of a process) for the purpose of reassigning the person's sex by changing physiological or other attributes of sex.'(s12(15)).
[202] ibid s12(12).
[203] World Health Organisation, 'Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation' <https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation> accessed 14 April 2023.
Karla Dhungana Sainju, Huda Zaidi, Niti Mishra and Akosua Kuffour, 'Xenophobic Bullying and COVID-19: An Exploration Using Big Data and Qualitative Analysis' (2022) 19(8) International Journal of Environmental Research and Public Health 4824, 14.
Alexandra Maftei, Andrei-Corneliu Holman, Ioan-Alex Merlici, 'Using fake news as means of cyber-bullying The link with compulsive internet use and online more disengagement' (2022) 127 Computers in Human Behaviour <https://www.sciencedirect.com/science/article/abs/pii/S0747563221003551> accessed 13 June 2023.

in a 12-month period,[204] though the majority do not make use of deeper controls.[205] If the usage of user empowerment features is similar, it is hard to predict how effective they may be, regardless of the fact its effect is only incidental.

This aspect of the new approach to harmful content under the Bill has a very limited possibility of addressing anti-information. It may also work against the marketplace of ideas by limiting alternative ideas the idea-consumer may encounter. User empowerment features could strengthen the echo chambers that currently exist and prevent alternative ideas replacing anti-information that users have come to believe.

Chapter 4.2.3: False Communications Offence

The most direct means for addressing anti-information within the Bill is the false communications offence. The offence originates from a Law Commission report,[206] and the Government accepted the recommendation of the Joint Committee for it form part of the Bill.[207] The offence is committed if a person (a) 'sends, transmits or publishes a communication,'[208] '(b) the message conveys information that the person knows to be false, (c) at the time of sending it, the person intended the message, or the information in it, to cause non-trivial psychological or physical harm to a likely audience, and (d) the person has no reasonable excuse for sending the message.'[209] "Likely audience" includes individuals that it is reasonably foreseeable would encounter the message,[210] or encounter a subsequent message that forwards it or shares its content.[211] The commission of the offence may result in a sentence on

[204] Digital Information World, 'Nearly 80% of Social Media Users have Adjusted their Privacy Settings in the Last Year' <https://www.digitalinformationworld.com/2019/10/research-shows-internet-users-taking-action-on-privacy.html#:~:text=Around%2079.2%20percent%20of%20the,profiles%20due%20to%20privacy%20concerns.> accessed 15 April 2023.
[205] Statista, 'Steps taken by global internet users to protect online activities and personal information as of December 2022' <https://www.statista.com/statistics/617422/online-privacy-measures-worldwide/> accessed 15 April 2023.
[206] Law Commission, *Modernising Communications Offences* (Law Com No 399, 2021) para 3.14.
[207] Department for Digital, Culture, Media & Sport, *Government Response to the Report of the Joint Committee on the Draft Online Safety Bill* (CP640, 2022) paras 93-95 cf Joint Committee on the Draft Online Safety Bill, *Draft Online Safety Bill* (2021-22, HL 129, HC 609) paras 131 and 135.
[208] Online Safety Bill (n 135) s160(1)(a) cf s163(2)(a).
[209] ibid s160(1).
[210] '"Encounter", in relation to a message, means read, view, hear or otherwise experience the message.' – Online Safety Bill (n 135) s163(5).
[211] Online Safety Bill (n 135) s160(2).

summary conviction of at most 6 months imprisonment, a fine, or both.[212] The requirement that the sender knows the information to be false means that it only applies to disinformation.[213]

The Law Commission's suggestion was intended to safeguard political expression through the requirement of intention as to harm and "no reasonable excuse,"[214] which would require the court to consider whether 'a communication was or was intended as a contribution to a matter of public interest.'[215] This was recognised by the Joint Committee that said 'It is also unclear whether the [false communication] offence would assist in cases of disinformation trying to disrupt elections, as the harm is based on psychological or physical harm, rather than harm to an institution, process, state or society.'[216] The Joint Committee's recommendation that this be addressed in the then Elections Bill was then ignored.[217]

The offence therefore has a limited scope which applies only to disinformation and specifically that which is intended to harm the audience physically or psychologically. Therefore, it may not cover health disinformation such as claiming ivermectin is a Covid-19 treatment,[218] as this would require stretching the definition of 'knows to be false' to those that genuinely believed the information but should have known it to be false.[219] Nonetheless, it would capture deceptive behaviour such as promoting "toxic" fitness regimes like "hormone balancing smoothies."[220]

---

[212] ibid s160(5)-(6).
[213] Law Commission, *Modernising Communication Offences* (n 206) para 3.30.
[214] ibid paras 2.10 and 3.67.
[215] ibid para 2.152.
[216] Joint Committee on the Draft Online Safety Bill (n 117) para 100.
[217] ibid cf DDCMS (n 207) paras 74-79, and Elections Act 2022.
[218] See for example 'Effective prevention and treatment for all respiratory viruses including Covid and Influence' (Doctor Myhill, July 2022) <https://www.drmyhill.co.uk/wiki/Effective_prevention_and_treatment_for_all_respiratory_viruses_including_Covid_and_Influenza> accessed 23 April 2022 cf Emily Cleary, 'GP suspended after pushing vitamins and ivermectin to treat COVID' (Yahoo!News, 6 February 2023) <https://uk.news.yahoo.com/gp-suspended-after-pushing-vitamins-and-ivermectin-to-treat-covid-172806333.html> accessed 23 April 2023.
[219] For example the GP that pushes a false treatment, ibid.
[220] Josie O'Brien, 'ABSOLUTE LIES: I was a toxic fitness influencer – here's the lies I told and why you shouldn't buy into ab workouts for a flat belly' (The Sun, 4 January 2023) <https://www.thesun.co.uk/fabulous/20936909/toxic-fitness-influencer-abs-lies/> accessed 23 April 2023.

All services have duties to perform risk assessments regarding illegal content,[221] and user-to-user services must 'swiftly take down' any illegal content,[222] and search services must minimise the risk of individuals encountering illegal content.[223] Illegal content is that which 'amounts to a relevant offence,'[224] which includes offences created by the Online Safety Bill.[225] Therefore, services will have a duty to remove content that in their view amounts to the false communication offence.

Although applying asymmetrically to only some forms of anti-information, this offence represents a step towards a preventative strategy, which is the preferred strategy. A preventative strategy is particularly effective because many of anti-information's harms may be caused when it is first encountered.[226]

The offence may be of limited use because of its high threshold. A requirement of intent on the speaker to cause harm means the speaker not only has to be deceptive but also to act with malice. This means that those spreading anti-information for a benign purpose, such as their own amusement,[227] would not commit an offence regardless of the harm they cause. A lesser level of intent being required for large scale harm – that the speaker should have anticipated – may be appropriate. Such an approach could be appropriate for the societal harms which the offence currently omits.[228] This would help to address the role of influence in the dissemination of anti-information. An individual with a large platform that amplifies content should anticipate the harm that could result should that information be false.[229]

[221] Online Safety Bill (n 135) s8 and s22.
[222] ibid s9(3).
[223] ibid s23(3).
[224] ibid s53(2).
[225] ibid s53(4)(b) cf s53(5)(c).
[226] See Chapter 2.2.
[227] Erin Buckels, Paul Trapnell, and Delroy Paulhus, 'Trolls just want to have fun' (2014) 67 Personality and Individual Differences 97-102.
[228] Joint Committee on the Draft Online Safety Bill (n 117) para 34-5.
[229] Joint Committee on the Draft Online Safety Bill (n 117) para 104.

Chapter 4.2.4: Appropriateness of the Online Safety Bill's Approach

Terms of Service and Private Censorship Concerns

A reliance upon platforms to adopt measures raises issues of private censorship, particularly when addressing an issue through a platform's TOS.

TOS may be changed at any point in time,[230] and a user often has not encountered the TOS since they first registered to the service.[231] TOS are ultimately unclear to users and are often a source of frustration when action is taken against them,[232] despite efforts to 'translate' the TOS into layman's terms.[233]

Additionally, since enforcing the TOS often relies upon other users flagging content they believe might violate the TOS,[234] the rules are not necessarily consistently applied.[235] The role that users play in flagging content is certainly critical but not clear.[236]

This is particularly concerning given that users may be censored through "mass reporting," where many accounts report one person or one particular post in an effort to get them or their content removed from the site. Although this is not necessarily a concern where the speaker has breached the TOS, it often appears to result in unjustified bans or suspensions.[237]

---

[230] Ethan Zuckerman, 'Intermediary Censorship' in Ronald Deibert, John Palfrey, Rafal Rohozinski and Jonathan Zittrain, *Access Controlled: The Shaping of Power, Rights and Rule in Cyberspace* (The MIT Press 2010) 75. Also, Sarah Myers West, 'Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms' (2018) 20(11) New Media & Society 4366-4383, 4369.
[231] West (n 231).
[232] ibid 4380. Also, Brian Contreras, '"I need my girlfriend off TikTok": How hackers game abuse-reporting systems' (Los Angeles Times, 3 December 2021) <https://www.latimes.com/business/technology/story/2021-12-03/inside-tiktoks-mass-reporting-problem> accessed 2 May 2023.
[233] West (n 231) 4370.
[234] ibid 4373-4. Also, Jessica Feezell, Meredith Conroy, Barbara Gomez-Aguinaga and John Wagner, 'Who Gets Flagged? An Experiment On Censorship and Bias in Social Media Reporting' (2023) 56(2) Political Science & Politics 222-26.
[235] Alex Dalbey, 'Trans people keep getting suspended from Twitter–and they want answers (updated)' (daily dot, 21 May 2021) <https://www.dailydot.com/irl/trans-twitter-bans/> accessed 2 May 2023.
[236] Feezell et al. (n 235) 223-5.
[237] Contreras (n 233) and Dalbey (n 236). Also, Sawdah Bhaimiya, 'Several left-wing activists had their Twitter accounts suspended after a false-report campaign by far-right users' (INSIDER, 1 December 2022) <https://www.businessinsider.com/left-wing-activists-banned-from-twitter-after-false-report-2022-11?r=US&IR=T> accessed 2 May 2023.

Online Safety Bill's Overall Appropriateness for Addressing the Anti-Information Issue

Ultimately, with regards to anti-information the Online Safety Bill does little to move beyond industry self-regulation, particularly due to the heavy reliance on the TOS and transparency duties. Whilst, the false communication offence does represent a step towards a preventative approach, it is limited in its applicability. Should the preventative approach that the false communication offence represents be strengthened so as to address the societal harms that it currently omits, the Online Safety Bill would be a desirable piece of legislation with regards to anti-information.

At present, it may be effective at addressing the role of influence and wealth, but this will ultimately depend on service providers and Ofcom's enforcement. Should service providers address anti-information through their TOS, and their TOS duties are complied with and enforced effectively otherwise, it could address the role of wealth and influence. The user empowerment duties pose a risk of reinforcing the echo chambers that are currently common on social media however, this may not be an issue if anti-information is effectively dealt with by service providers and the idea-consumer does not encounter it, and instead encounters ideas that replace the anti-information they have come to believe outside of the platform. Such limitations could be rectified by Ofcom, particularly through requiring platforms to take certain steps when they find that the platform has failed in their duty.

Ultimately, the likely effectiveness of the Bill is uncertain. A move towards a stronger preventative approach and more specific requirements for how platforms should form policies to address anti-information would be welcomed. Should the Bill be enacted in its current form, Ofcom's enforcement of the TOS duties with regards to anti-information, assuming service providers add it to their TOS, would be critical to effectively addressing anti-information.

Chapter 4.3: Directly Addressing Anti-Information

The possible extended imprinting requirements and Online Safety Bill have uncertain effectiveness and depend upon effective enforcement. Even if effectively enforced however, they suffer a few limitations. Principally, they do not provide alternative ideas for the idea-consumer to replace their belief in anti-

information with. Whatever they do to encourage false ideas aren't believed, they don't encourage truthful ideas to replace them. Additionally, other than the TOS duties which may not even apply to anti-information, each is limited to certain forms of anti-information.

To the extent that preventative measures appear to be developing, there is positive development of an effective anti-information strategy. However, preventative and removal/access limitation measures should be supported with information correction measures, so that when anti-information is refuted for the idea-consumer, there are alternative ideas for them to believe, limiting the likelihood they continue to foster belief in false information and potentially preventing their belief being reinforced by psychological effects such as confirmation bias.

Chapter 5: Bot Speech

It was demonstrated in Chapter 2 that bot technology is capable of mass dissemination of anti-information. Such dissemination may overwhelm the idea-consumer, and if made at opportune time, such as before an election, dissemination of anti-information through bots could be extremely effective at convincing the idea-consumer of the truth of many ideas.

This Chapter first discusses California's bot disclosure law. The law requires bots that speak to Californians to identify themselves as bots. Although it will be concluded that the effectiveness of bot identity disclosure is questionable, there is a distinct lack of proposals for regulating bot speech in the UK and therefore, it is appropriate to consider whether a measure similar to California's identity law could be ECHR compliant and therefore, if it is an option available to the UK.

Before considering ECHR compliance, this chapter will discuss speech interests particular to bot speech. This discussion will include a general speech interest in the variety of bot speech, before discussing three interests identified through criticism of California's bot disclosure law with regards to the First Amendment. These First Amendment concerns would likely manifest differently under the Convention system, however, it is worth considering what impact they may have.

This Chapter concludes bot identity transparency is a weak solution to the anti-information issue in relation to bot speech. Given that such a measure represents a "more speech" solution, this Chapter finds that improving transparency does not prevent the idea-consumer from being overwhelmed by bot speech. Although finding that a bot identity measure would be ECHR compliant, and therefore would be an option for the UK, this Chapter concludes that a stricter measure, such as removal of bot speech or prevention by criminalisation would be more appropriate.

Chapter 5.1: California's Bot Disclosure Law

"SB-1001 Bots: disclosure" was approved by the Governor and chaptered by the Secretary of State on 28 September 2018. SB-1001 adds a chapter comprised of 4 sections to California's Business and Professions Code. Section 17940 of that Code defines "bot", "online", "online platform" and "person" for the purposes of the Chapter, whilst Section 17941 details the offence and Sections 17942 and 17943 contain the usual severability clause, commencement and other expected aspects.

The unlawful act is detailed as follows:

> Section 17941(a):
>
> 'It shall be unlawful for any person to use a bot to communicate or interact with another person in California online, with the intent to mislead the other person about its artificial identity for the purpose of knowingly deceiving the person about the content of the communication in order to incentivize a purchase or sale of goods or services in a commercial transaction or to influence a vote in an election. A person using a bot shall not be liable under this section if the person discloses that it is a bot.'
>
> Section 17941(b):
>
> 'The disclosure required by this section shall be *clear, conspicuous, and reasonably designed to inform* persons with whom the bot communicates or interacts that it is a bot.'

The unlawful act is simple – a person[1] may not communicate or interact with another person[2] to advertise goods and services, or to influence a vote, whilst not disclosing the artificial identity of the

[1] Meaning, 'a natural person, corporation, limited liability company, partnership, joint venture, association, estate, trust, government, governmental subdivision or agency, or other legal entity or any combination thereof.' SB-1001 Bots: disclosure, cf Business and Professions Code, Part 3 of Division 7, Chapter 6, Section 17940(d) (California). Full text of SB-1001 available here: <https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1001> accessed 1 January 2022. Henceforth, SB-1001.
Relevant Chapter of the Business and Professions Code available here: <https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=BPC&division=7.&title=&part=3.&chapter=6.&article=> accessed 1 January 2022.

[2] Same meaning, ibid.

"bot" through which the communication or interaction takes place. 'Call it Isaac Asimov's fourth Rule of Robotics: A robot may not pretend to be a human being.'[3]

Phrasing that can also be found in the Honest Ads Bill in Section 17941(b) has been used. The Honest Ads Act would require disclosure to be stated in a 'clear and conspicuous manner'[4] but the wording of SB-1001 goes further and adds that disclosure must be "reasonably designed to inform" people that it is a bot. This is likely intended to have the same effect as a provision in the Honest Ads Act that says a statement is not clear and conspicuous 'if it is difficult to read or hear or if the placement is easily overlooked.'[5] Therefore, a disclosure that took the form of, for example, hard-to-read small print at the end of an ad, or in this case, at the termination of communication (after having influenced the other person), would not meet the requirements of SB-1001.

SB-1001 defines a "bot" as 'an automated online account where all or substantially all of the actions or posts of that account are not the result of a person.'[6] This definition is suitable for catching chatbots and social bots,[7] therefore the bots that are most prevalent at spreading anti-information would therefore be caught by the Act.[8]

SB-1001's definition appears to be effective for the Act's intended purpose, and is unlikely to limit the Act's effectiveness. However, the effect of transparency as to the speaker's bot identity may not be effective at mitigating the harms SB-1001 aims to prevent.

[3] Barry Stricke, 'People v. Robots: A Roadmap for Enforcing California's New Online Bot Disclosure Act' (2020) 22 Vanderbilt Journal of Entertainment and Technology Law 839.
[4] The text is the same across all three introductions of the Bill.
Text of 2017 Senate Bill, <https://www.congress.gov/bill/115th-congress/senate-bill/1989/text> accessed 15 January 2022.
Text of 2019 Senate Bill, <https://www.congress.gov/bill/116th-congress/senate-bill/1356/text> accessed 15 January 2022.
Text of 2019 House of Representatives Bill, <https://www.congress.gov/bill/116th-congress/house-bill/2592/text> accessed 15 January 2022.
Henceforth, the Honest Ads Bill. See Section 7(a)(1).
[5] Honest Ads Act (n 4) Section 9(a)(2).
[6] SB-1001 (n 1) cf. Business and Professions Code (n 1) s17940(a).
[7] Stricke (n 3) 850-3.
[8] See PR Chamberlain, 'Twitter as a Vector for Disinformation' (2010) 9(1) Journal of Information Warfare 11, 11.

Chapter 5.1.1: Effectiveness and Bot Identity Transparency

Despite the headway that SB-1001 makes, transparency alone is insufficient to protect against the threat of bots.[9] The effect of SB-1001 will be dependent on the extent to which transparency as to the bot's artificial identity will prevent the harms to which it is addressed.

In commercial interactions, bot identity transparency appears to make a substantial impact. There is a preference amongst consumers to interact with a human over a bot.[10] Where the bot identity is disclosed, a significant reduction in purchase rates was observed because the perception was that a bot would be less knowledgeable and less empathetic.[11]

There is limited research regarding the role of bot identity transparency plays with regards to bots that aim to influence elections. However, there is research dedicated to examining the effects of a bot's "humanness."[12]

Uncanny valley feelings – discomfort at almost human objects – has been observed in sophisticated bots where users knew they were speaking to a bot,[13] however these feelings were contrarily observed alongside feelings of trust and comfort.[14] In another study, such feelings did not arise when users did not know they were talking to a bot,[15] however, this may have been due to lack of sophistication and therefore, perceived humanness of the bot.[16]

[9] The limitations of transparency-improving measures generally were discussed in Chapter 3.1.
[10] Roberta De Cicco, Susanna Cristina Lima da Costa e Silva and Riccardo Palumbo, 'Should a Chatbot Disclose Itself? Implications for an Online Conversational Retailer' (2021) Conversations 2020, available at <https://link.springer.com/chapter/10.1007/978-3-030-68288-0_1> accessed 27 April 2023
[11] Xeuming Lou, Siliang Tong, Zheng Fang and Zhe Qu 'Fontiers Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases' (2019) 38(6) Marketing Science 937-947, 945.
[12] Amon Rapp, Lorenzo Curti and Arianna Boldi, 'The human side of human-chatbot interaction: A systemic literature review of ten years of research on text-based chatbots' (2021) 151 International Journal of Human-Computer Studies, available at <https://www.sciencedirect.com/science/article/pii/S1071581921000483> accessed 27 April 2023.
[13] Vivian Ta, Caroline Griffith, Carolynn Boatfield, Xinyu Wang, Maria Civitello, Haley Bader, Esther DeCero and Alexia Loggarakis, 'User Experiences of Social Support From Companion Chatbots in Everyday Contexts:: Thematic Analysis' (2020) 22(3) Journal of Medical Internet Research, available at <https://www.jmir.org/2020/3/e16235/> accessed 27 April 2023.
[14] ibid.
[15] Marita Skjuve, Ida Maria Haugstveit, Asbjørn Følstad and Petter Bae Brandtzaeg, 'Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human–chatbot interaction' (2019) 15(1) Human Technology 30-54, 47.
[16] ibid.

Regardless, bots are capable of eliciting trust from users and this is likely to grow as bots become more sophisticated.[17] Additionally, there is evidence that the development of bonds between user and bot, and feelings of trust are encouraged by greater perceptions of the humanness of the bot.[18] Whereas users that know of a bot's artificial identity are less likely to make effort to communicate effectively with it.[19] Transparency as to the bot's identity would presumably interfere with its ability to communicate effectively with the user and inflict the harms that SB-1001 attempts to protect against.

One study found that users were more likely to make a charitable donation when they believed they were talking to a human than a bot,[20] however, the same study also observed users that did not believe they were talking to a bot despite disclosure as to its artificial identity because 'the bot responses "were appropriate and heartfelt."'[21]

Bot identity disclosure may play a substantial role in preventing the harms SB-1001 addresses, however, its impact is somewhat unclear. To make effective use of this form of transparency, legislators that seek to adopt a similar strategy should refer to the latest research and human-bot interaction experts.

Chapter 5.2: Relevant Speech Interests in Regulating Bot Speech

This section examines four speech interests relating to the regulation of bot speech. It finds that the wide variety of bot speech means that the application of a general measure to bot speech without unjustifiably interfering with speech interests is not possible. Further, that the issue of "compelled

[17] Asbjørn Følstad, Cecilie Bertinussen Nordheim & Cato Alexander Bjørkli, 'What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study' (2018) Internet Science, available at <https://link.springer.com/chapter/10.1007/978-3-030-01437-7_16> accessed 27 April 2023.
[18] ibid. See also, Kien Hoa Ly, Ann-Marie Ly and Gerhard Andersson, 'A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods' (2017) 10 Internet Interventions 39-46, 45-46.
[19] Kevin Corti and Alex Gillespie, 'Co-constructing intersubjectivity with artificial conversational agents: People are more likely to initiate repairs of misunderstandings with agents represented as human' (2016) 58 Computers in Human Behavior 431-442, 432 and 440-1.
[20] Weiyan Shi, Xeuwei Wang, Yoo Jung Oh, Jingwen Zhang, Saurav Sahay and Zhou Yu, 'Effects of Persuasive Dialogues: Testing Bot Identities and Inquiry Strategies' (2020) Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems 1-13, 10.
[21] ibid 7.

speech" does not apply to bot identity disclosure because such a principle is limited to protected ideas, not factual information. The religious baker may believe that baking a cake that supports gay marriage is a sin and supporting gay marriage is incorrect, but the speaker communicating through a bot does not believe that it is not a bot. Additionally, this section concludes that given that bot identity disclosure does not amount to suppressing a particular viewpoint and that "content-neutrality" is not a requirement under Art.10, ECHR, there should be no issue faced by bot identity disclosure in the Convention system. Finally, this section also finds that whilst concerns regarding "unmasking" the anonymous speaker could potentially be avoided, Art.8, ECHR concerning the right to privacy would play a significant role in determining how such an interest may manifest.

Chapter 5.2.1: (Unique) Expression through Bot Speech

Importantly, the law must respond to emerging forms of speech appropriately,[22] and in the realm of social media for example, it is clear governments are expected to protect users from harm without unjustly interfering with freedom of expression.[23] Bots however would not be able to receive protection under Art.10 independently of their speakers, though the right for the speaker to express themselves through a bot is likely to be protected. The variety of speech bots enable means a blanket policy towards bot speech would likely unjustly interfere with freedom of expression.

A concern is that the complexity of issues surrounding enforcement of bot disclosure laws may lead private actors and legislatures to censor "bot speech" to avoid the issue entirely.[24] This is a concern because to consider all bot speech as a single category masks the wide variety of bot speech that exists, including bots that exist for artistic purposes, which is in and of itself a broad category. Some examples from Twitter include: bots that share art of others (as specific as the art of individuals, including that of

[22] See *Spiller v Joseph* [2010] UKSC 53, [99] and [131]. Also, Office of the United Nations High Commissioner for Human Rights, 'New and emerging technologies need urgent oversight and robust transparency: UN experts' (2 June 2023) <https://www.ohchr.org/en/press-releases/2023/06/new-and-emerging-technologies-need-urgent-oversight-and-robust-transparency> accessed 3 June 2023.
[23] Recommendation CM/Rec(2012)4 of the Committee of Ministers to member states on the protection of human rights with regard to social networking services (4 April 2012) para 6.
[24] Madeline Lamo & Ryan Calo, 'Regulating Bot Speech' (2019) 66(4) UCLA Law Review 988-1028, 1023-5.

American poet, Emily Dickinson,[25] and as general as just 'beautiful paintings'[26]), bots which generate art abstractly (for example 'micropoetry' bot @poem_exe,[27] or @greatartbot[28]) and bots which generate art that reflects society (for example @censusAmericans which generates a profile describing an American using census data,[29] and @oliviataters[30] a chatbot 'whose feed reads like the unedited thoughts of the British teenager inside all of us'[31]). Lamo and Calo correctly caution against the suppression of a new form of speech entirely. Their request for bot regulation to be 'aimed at (1) particular categories of bots, (2) within specific contexts, and (3) supported by the specific harms the government hopes to mitigate'[32] is well-founded, and fulfilled by SB-1001.

Bot regulation must account for the different types of bot speech that can be encountered online and the harms they represent in order to adequately safeguard users and their freedom of expression. A measure that targets bots generally rather than for the harms they can cause would risk unjustifiable interfering with the freedom of expression of other, innocuous bot speakers.

Chapter 5.2.2: Compelled Speech

Bot disclosure as a form of compelled speech could be an issue for SB-1001 and other legislation like it. '[T]he choice to speak includes within it the choice of what not to say.'[33] Lamo and Calo highlight that the American jurisprudence 'suggest[s] that the government cannot force commercial speakers to endorse ideas contrary to their own interests'[34]

However, they also note an exception, whereby details such as nutritional details may have mandated disclosure because 'Such disclosure furthers, rather than hinders, the First Amendment goal of the discovery of truth and contributes to the efficiency of the "marketplace of ideas"' according to the

[25] @emilydicknsnbot, <https://twitter.com/emilydicknsnbot> accessed 25 January 2022.
[26] @artbot_ <https://twitter.com/artbot__> accessed 25 January 2022.
[27] @poem_exe <https://twitter.com/poem_exe> accessed 25 January 2022.
[28] @greatartbot <https://twitter.com/greatartbot> accessed 25 January 2022.
[29] @censusAmericans <https://twitter.com/censusamericans> accessed 25 January 2022.
[30] @oliviataters <https://twitter.com/oliviataters> accessed 25 January 2022.
[31] Peter Hess, 'The User Experience Researcher Who Was Fooled by a Twitter Bot' <https://www.vice.com/en/article/kb7bmn/olivia-taters-twitter-bots> accessed 25 January 2022.
[32] Lamo & Calo (n 23) 1027.
[33] *Pacific Gas & Electric Co. v Public Utilities Commission* [1986] 475 US 1, 16 (United States of America).
[34] Lamo and Calo (n 23) 1012.

Second Circuit Court of Appeal.[35] This disclosure aids the marketplace of ideas because it's "accurate", "factual" and "truthful."[36]

The same could be said of laws that require a bot to disclose that it is a bot. This disclosure reveals little information, and it does not mean that bots cannot be used to sell products and influence elections, it just means their potential for deception is limited (e.g., social bot networks may be less convincing). Limiting the potential for deception must similarly contribute to the effective functioning of the marketplace of ideas.

Although principles relating to compelled speech could be an issue for SB-1001, its contribution to the marketplace of ideas would likely be acknowledged and similarly received as disclosure requirements in advertising and commerce.

Although there is not a robust doctrine of compelled speech under UK law or Strasbourg jurisprudence. Though the ECtHR has not ruled out the existence of a "negative right" to not be compelled to speak, it insists that the issue is to be considered on a case-by-case basis.[37] With the decision in *Lee v Ashers,*[38] steps towards developing a domestic doctrine have been taken.[39]

The very first step however could be attributed to Lord Dyson's analysis in *RT (Zimbabwe) v Secretary of State for the Home Department*.[40] By analogy of the right to not hold religious beliefs enshrined under Art.9 of the Convention,[41] Dyson found there to be 'no basis in principle for treating the right to hold and not hold political beliefs differently.'[42] His summary: 'Nobody should be forced to have or express a political opinion in which he does not believe.'[43]

---

[35] *New York State Restaurant Association v New York City Board of Heath* [2009] 556 F3d 114, 132 (2d Cir) (United States of America) quoting *National Electrical Manufacturers Association v Sorrell* [2001] 272 F3d 104 (2d Cir), 113-14 (United States of America).
[36] ibid.
[37] *Gillberg v Sweden* App. No.41723/06, [86]; recently applied in *Wanner v Germany* App. No.26892/12, [39].
[38] *Lee v Ashers Baking Company Ltd and Others (Northern Ireland)* [2018] UKSC 49.
[39] Jacob Rowbottom, 'Cakes, Gay Marriage and the Right against Compelled Speech' (UK Constitutional Law Association, 16 October 2018) <https://ukconstitutionallaw.org/2018/10/16/jacob-rowbottom-cakes-gay-marriage-and-the-right-against-compelled-speech/> accessed 18 October 2022.
[40] *RT (Zimbabwe)* [2012] UKSC 38.
[41] *Buscarini and Others v San Marino* App. No.24645/94, [34]. Cf. *RT (Zimbabwe)* (n 40) [35].
[42] *RT (Zimbabwe)* (n 39) [36].
[43] ibid [42].

Lady Hale's judgment in *Lee v Ashers* followed the same logic.[44] The case concerned a bakery which refused to make a cake with a message that supported same-sex marriage due to their religious beliefs. After discussion of Articles 9 and 10 of the Convention, and concurring with Dyson's view in *RT* that they support a right to not express an opinion or belief,[45] she found that 'they would be entitled to refuse to do that whatever the message conveyed by the icing on the cake'[46] and immediately followed that with examples limited to religious and political expression, 'support for living in sin, support for a particular political party, support for a particular religious denomination.'[47]

Should the development of a "compelled speech"[48] doctrine be principally limited to speech that is religious or political,[49] it would pose no issue to a version of SB-1001 in the UK, as the speech it compels is factual and objective.[50] Whilst development of the right not to speak may have implications for many forms of expression, such as a teacher who is compelled to teach facts that they disagree with or a copy-editor working on a piece they do not support,[51] compelling a person to ensure their bot declares itself to be a bot is more analogous to requiring ingredients to be listed on food, health warnings on cigarettes, or that a political advertisement was paid for by a certain group. The speaker does not believe that the bot is not a bot, unlike the conscientious objector who does not believe in a certain religious or political doctrine.

Chapter 5.2.3: Content-neutrality

Content-neutrality is required by the First Amendment. If a measure is content-based, it 'applies to particular speech because of the topic discussed or the idea or message expressed'[52] and 'is subject to strict scrutiny regardless of the government's benign motive, content-neutral justification, or lack of "animus towards the ideas contained" in the regulated speech.'[53]

[44] *Lee v Ashers* (n 37) [49]-[55].
[45] ibid.
[46] ibid [55].
[47] ibid.
[48] Jacob Rowbottom (n 38) cf. *Lee v Ashers* (n 37) [53].
[49] Jacob Rowbottom (n 38).
[50] See Chapter 2.1.
[51] Jacob Rowbottom (n 38).
[52] *Reed v Town of Gilbert, Arizona* [2015] 576 US 155, 163 (USA).
[53] ibid 156 and 165.

Weaver argues that by targeting 'communication in order to incentivize a purchase or sale of goods or services in a commercial transaction of to influence a vote in an election' SB-1001 "moves away" from being content-neutral since it targets a specific type of communication.[54] Hines is of a different opinion; since SB-1001 'does not directly discriminate against the viewpoints of the bot user'[55] it is content-neutral.

Since SB-1001 does not regulate any specific viewpoint, it 'does not suppress a bot user's viewpoint'[56] unlike a "gag order" for example.[57] The effect of the Act is therefore to reduce the chance a user interacting with a bot can be misled because although the bots' view is unaltered and not interfered with, the way it is presented puts the users in a more knowledgeable position.[58] The effects of SB-1001 are content-neutral as the interference with the First Amendment 'does not depend upon the content.'[59]

Although Strasbourg jurisprudence does not categorise legislation as content-based or content-neutral, there is sufficient case law on related matters that suggest the approach it may take with regards to bot speech, including cases relating to viewpoint suppression and censorship generally.

In *Bayev and Others v Russia*, the ECtHR rejected arguments that measures adopted to outlaw "the promotion of homosexuality" could be justified on grounds of morals, health or protection of the rights of others and found a violation of Art.10 in conjunction with Art.14, Prohibition of discrimination.[60] Suppression of such a view was found not to have a legitimate aim as the adoption of such a measure 'is incompatible with the notions of equality, pluralism and tolerance inherent in a democratic society.'[61] Unlike under the First Amendment, viewpoint suppression can be justified in a democratic society, given that it has a legitimate aim under Art.10(2).

---

[54] John Frank Weaver, 'Everything Is Not Terminator: We Need the California Bot Bill But We Need It to Be Better' (2018) 1(6) The Journal of Robotics, Artificial Intelligence & Law 431, 433.
[55] Matthew Hines, 'I Smell a Bot: California's SB 1001, Free Speech, and the Future of Bot Regulation' (2019) 57 Houston Law Review 405, 426.
[56] ibid 425-6.
[57] Example given by Hines (n 55) 426, *Doe v Gonzales* [2005] 386 FSupp 2d 66, 75 (D. Conn.) (United States of America).
[58] Hines (n 55) 426.
[59] *Turner Broadcasting System v Federal Communications Commission* [1994] 512 US 622, 643-44 (United States of America) cf Hines (n 55) 426, at note 139-40.
[60] App. Nos.67667/09, 44092/12, 56717/12, [65]-[85].
[61] ibid [83].

Similarly to the First Amendment's content-based jurisprudence, the ECtHR has found that suppressing particular ideas/information to have a chilling effect. In *Vajnai v Hungary*,[62] the suppression of a red star symbol by Hungary's criminal code[63] resulted in a violation of Art.10. The ECtHR found the ban to be too broad and the uncertainty therefore created by it to entail 'a chilling effect on freedom of expression and self-censorship.'[64] The mere existence of such a measure has a chilling effect in forcing the right-holder 'to modify [their] conduct by displaying self-restraint… in order not to risk prosecution.'[65] The measures in *Vajnai*, however, concern a particular expression, as opposed to the mode by which it is communicated. Although protection of means of transmission does fall within Art.10.[66]

With regards to censoring measures, the ECtHR has found violations where interference has prevented criticism of the government,[67] where future criticism of the government has been deterred,[68] and where the interference "hampers" the purveying of information which aids discussion,[69] including the provision of alternative viewpoints.[70] This applies beyond the public watchdog,[71] i.e., the press, NGOs, and academics,[72] to the freedom of expression of ordinary persons.[73]

There is no particular viewpoint that would be suppressed by bot identity disclosure. Further, regardless of whether content-neutrality requires the type of expression to also not be interfered with, content-neutrality is not a requirement of Art.10 and interference with bot speech under it could be justified by reference to its harms.

---

[62] App. No.33629/06.
[63] ibid [15].
[64] ibid [54].
[65] *Altuğ Taner Akçam v Turkey* App. No.27520/07, [75].
[66] *Magyar Kétfarkú Kutya Párt v Hungary* App. No.201/17, [87] and *Ahmet Yıldırım v Turkey* App. No.3111/10, [50].
[67] *Toranzo Gomez v Spain* App. No.26922/14, [64], *Lingens v Austria* App. No.9815/82, [44].
[68] *Lingens* ibid, *Lewandowska-Malec v Poland* App. No.39660/07, [70], *Monnat v Switzerland* App. No.73604/01, [70].
[69] *Lingens* ibid, *Monnat* ibid, *Barthold v Germany* App. No.8734/79, [58].
[70] *Ali Gürbüz v Turkey* App. No.52497/08, [77].
[71] The public watchdog role is usually seen as being performed by the press, e.g., *Bladet Tromsø and Stensaas v. Norway* App. No.21980/93, [59].
[72] However, the public watchdog does extend further, *Animal Defenders International v UK* App. No.48876/08, [103] cf. *Magyar Helsinki Bizottság v Hungary* App. No.18030/11, [166]-[168].
[73] In *Barthold v Germany* (n 69) the ECtHR considers these factors for 'members of the liberal professions.'

Chapter 5.2.4: Anonymous Speech

Protection of anonymous speech under the First Amendment has led to the development of a right to anonymity in First Amendment jurisprudence.

In *Talley v California*, it was said that identification requirements of the speaker 'would tend to restrict freedom to distribute information and thereby freedom of expression'[74] and that anonymous speech has 'played an important role in the progress of mankind.'[75] In doing so, the Supreme Court cited the 'obnoxious press licensing law of England... enforced on the Colonies... due in part to the knowledge that exposure of the names of printers, writers and distributors would lessen the circulation of literature critical of the government.'[76] The Supreme Court sees anonymous speech as an important part of the American political and constitutional tradition.[77]

Despite First Amendment challenges, disclosure requirements have been upheld. The disclosure requirements in the Federal Election Campaign Act of 1971 were upheld in *Buckley v Valeo*,[78] where the Supreme Court listed a number of government interests pursued by the requirements, including 'providing the electorate with information... to aid the voters in evaluating [candidates]... more precisely than is often possible solely on the basis of party labels and campaign speeches'[79] inter alia. To "open the basic processes of the federal election system" was seen by the Supreme Court as being 'a reasonable and minimally restrictive method of furthering First Amendment values'[80] and such a position reflects the "more speech" approach of the counterspeech doctrine.

Online, American courts appear to be applying the right to anonymity with as robust protection as in the real world; online, the right to anonymity remains strong yet limited.[81]

---

[74] *Talley v State of California* [1960] 362 US 60, 64 (United States of America).
[75] ibid.
[76] ibid.
[77] *Watchtower Bible and Tract Society of New York v Village of Stratton* [2002] 536 US 150, 162-166 (United States of America).
McIntyre v Ohio Elections Commission [1995] 514 US 334, 357 (United States of America).
[78] (1976) 424 US 1, 81 (United States of America).
[79] ibid 66-7.
[80] ibid 82.
[81] Jeff Kosseff, *The United States of Anonymous: How the First Amendment Shaped Online Speech* (Cornell University Press 2022) 112-141.

Stricke believes that SB-1001 would survive constitutional challenge and that it would not collide harshly with the right to anonymity since it 'only requires bots to disclose that they are not in fact a human, with no other condition modifying or otherwise characterising the speech, much less identifying the speaker.'[82] However, the risk to the right to anonymity is of great concern to Lamo and Calo. Their concern hinges on their finding that the exceptions to the right to anonymous speech are narrow,[83] and since there is no official means by which a person can reveal they are a person without revealing who they are, 'we must assume virtually every instance of enforcement will involve unmasking'.[84] Therefore, it appears unlikely bot disclosure laws could be enforced without guaranteeing infringement with the right to anonymous speech.[85]

Weaver argues that mandating bot self-identification 'is akin to requiring natural person speakers to identify themselves.'[86] This may be because Weaver understands bot speech as falling within the ambit of the First Amendment protection since the First Amendment does not mention humans or natural persons,[87] however, to collate bots and natural persons in such a way is to imbue bots with a level of personhood that they do not, at least as of yet, possess. The true "speaker" is the bot's programmer,[88] not the bot. It is simply a mode of communicating.

It would be possible to protect the right to anonymity with technological development.[89] Lamo and Calo suggest using a system to confirm the humanness of the speaker, either through a social media site and therefore, not requiring the user to reveal their name to the site, or through a third party.[90] However, the problem lies in such means not yet existing or in that they are not officially sanctioned and therefore, no obligation to take the confirmation into account would exist for a prosecutor, who could continue to investigate up until anonymity is lost.[91]

[82] Stricke (n 3) 889.
[83] Lamo and Calo (n 23) 1024.
[84] ibid 1023.
[85] ibid 1024.
[86] Weaver (n 54) 432.
[87] ibid 431.
[88] Hines (n 55).
[89] Lamo and Calo (n 23) 1023.
[90] Lamo and Calo (n 23) 1023.
[91] ibid 1023-4.

Therefore, on the basis of Lamo and Calo's analysis, to survive a free speech challenge, a bot disclosure law would need to carve another exception to the right to anonymity out of the jurisprudence. Turning to the case law on the right to anonymity regarding electoral speech,[92] the governmental interest of protecting electoral integrity has been considered sufficient by the Supreme Court to allow the unmasking of previously anonymous referendum petition signatories.[93]

The right to anonymity serves the greatest challenge in First Amendment jurisprudence to a law like SB-1001 because of the strength of protection afforded to anonymous speakers.

The requirement that a bot identifies itself could arguably avoid any anonymity concern, as to equate the bot to a natural person holding a right to anonymity is unpersuasive. Additionally, although the right to anonymity covers many forms of expression, there is no indication that it is a right to an anonymous means of expression. The speaker's anonymity is not interfered with by unmasking that they speak through a bot. However, an issue may arise where a speaker fails to have their bot disclose that it is a bot and they are resultingly unmasked in litigation.

Under Strasbourg's jurisprudence, there is a separate right to anonymity for expression beyond the right to privacy (Art.8).[94] The right is not 'absolute and must yield on occasion to other legitimate imperatives, such as the prevention of disorder or crime or the protection of the rights and freedoms of others.'[95]

The purpose of anonymity has been described similarly to the US Supreme Court's view: as a means to avoid 'reprisals or unwanted attention… [and promote] the free flow of ideas and information.'[96] Additionally the ECtHR has acknowledged that interfering with the right could 'deter [speakers] from contributing to debate and therefore lead to a chilling effect.'[97]

Nonetheless, an understanding of the requirements of anonymity with regards to speech would require paying particular attention to Art.8 as well as Art.10. The scope of this undertaking falls beyond the

[92] ibid 1021-2.
[93] *Doe v Reed* [2010] 561 US 186 (United States of America).
[94] See *Breyer v Germany* App. No.50001/12, [60]-[62].
[95] *KU v Finland* App. No.2872/02, [49].
[96] *Delfi AS v Estonia* App. No.64569/09, [147].
[97] *Standard Verlagsgesellschaft MBH v Austria (No 3)* App. No.39378/15, [74].

reach of this thesis. However, it is enough to identify that the speech interest in anonymous speech can be justifiably interfered with, as with other expression protected by Art.10.

Chapter 5.3: Bot Identity Transparency in the UK

There is no specific regulation of bot speech in the UK. If we consider it to be necessary to address the effects of new technology on the marketplace of ideas, particularly their role in spreading anti-information, then there appears to be a significant gap in the UK's approach to the issue.

The Online Safety Bill, in Part 3 (duties of care of regulated services) bots are included within the definition of (regulated) user-generated content: 'the reference to content generated, uploaded or shared by a user includes content generated, uploaded or shared by means of software or an automated tool applied by the user.'[98]

Lack of specific provision addressing the role of bots appears to be a significant gap therefore in the UK's approach to countering anti-information and generally, in regulating the idea market. Therefore, the compatibility with the ECHR of a bot identity transparency law like SB-1001 will be considered.

Chapter 5.3.1: Bots as Rights-Holders

There has been some debate over whether human rights should be recognised for artificial intelligence at some point in the future.[99] Regardless, bots are not within a category of persons capable of making an individual application to the European Court of Human Rights. The text of Article 34 of the Convention states:

> 'The Court may receive applications from any person, non-governmental organisation or group of individuals claiming to be the victim of a violation by one of the High Contracting Parties of the rights set forth in the Convention or the Protocols thereto. The High Contracting Parties undertake not to hinder in any way the effective exercise of this right.'

[98] Online Safety Bill HL Bill (2022-23) 87(Rev) s49(4)(a) cf s49(2)-(3).
[99] See John-Stewart Gordon and Ausrine Pasvenkiene, 'Human rights for robots? A literature review' (2021) 1(4) AI and Ethics 579-591.

A bot is unlikely to be able to form a non-governmental organisation without first being recognised as a person in national law, and "person" in Art.34 refers to natural persons as indicated by the French version of the Convention which uses the term, *toute personne physique*.[100] Therefore, bot speech may only be protected as a means through which natural persons speak – 'Article 10 applies not only to the content of information but also to the means of transmission or reception since any restriction imposed on the means necessarily interferes with the right to receive and impact information.'[101]

Chapter 5.3.2: Bot Identity Transparency as a Justifiable Interference

Given that bot speech is likely to be protected as a means through which a rights-bearer can exercise their art.10 rights, this section considers whether interference with bot speech by mandating identity transparency, similarly to SB-1001, is compliant with the ECHR.

A law like SB-1001 would meet the requirements of "prescribed by law" so long as it is accessible and its consequences are sufficiently foreseeable.[102]

In seeking to regulate against deception SB-1001 limits itself to bots that seek to influence an election or incentivise a commercial transaction. The legitimate aim pursued would differ depending on the objective of the bot. A state could pursue protection of the rights of others with regards to bots that attempt to deceive as to their identity in order to influence an election. The legitimate aim addresses 'the need to protect the electoral process as part of the democratic order,'[103] and could be applied preceding an election, or continuously.

Meanwhile a number of legitimate aims could be pursued depending on the particular commercial transaction that is incentivised. For example, the prevention of disorder or crime with regards to

[100] The ECHR follows the Vienna Convention on the Law of Treaties, which states in Art.33(3) that where a treaty is 'authenticated in two or more languages, the terms of a treaty are "presumed to have the same meaning in each authentic text."' – *Perinçek v Switzerland* App. No.27510/08, [149].
[101] *Autronic AG v Switzerland* App. No.12726/87, [47].
[102] *The Sunday Times v UK (No.1)*, App. No.6538/74, [49].
[103] *Animal Defenders International v UK* (n 72) [111].

fraudulent advertising and scammers,[104] or the protection of morals with regards to incentivised purchase of age-inappropriate applications or pornography.[105] However, this section will not consider the pursuit of these aims as the harms which a provision regarding incentivising commercial transactions appears to fall more squarely within fraudulent advertising,[106] or at least beyond anti-information and the scope of this thesis. However, a measure similar to SB-1001 that targets speech which aims to influence the result of an election could fall under the legitimate aim of protection of the rights of others, notably Article 3 of Protocol 1, the right to free elections.

The necessity of the interference could be justified. It is ultimately a very minimal interference on freedom of expression. It is unlikely there is a less restrictive measure which would achieve the same aim.[107] Whilst it could potentially be a disproportionate interference depending on the punishment, allegations of disproportionality should be easy to avoid. A low fine or even the possibility of an order being issued that requires the speaker to have their bots disclose their artificial identity, and to continue to comply with the future, strikes the right balance between their freedom of expression and the legitimate aim.

Although there were no concerns with regards to compelled or content-biased speech, anonymous speech could prove to be an issue for a measure like SB-1001 seeking Convention compliance. However, examining the scope of the right to anonymous speech requires an examination of Art.8 which is beyond the scope of this thesis.

Chapter 5.4: Appropriateness of a Bot Identity Transparency Approach

[104] Joint Committee on the Draft Online Safety Bill, *Draft Online Safety Bill* (2021-22, HL 129, HC 609) pp161-162, Appendix 1.
[105] ibid paras 214-9.
[106] As in the Online Safety Bill (n 98) s33-35.
[107] *Glor v Switzerland* App.No.13444/04, [94].

Although a measure like bot identity transparency could be welcomed if it is shown that bot identity transparency is effective,[108] it is ultimately a weak solution to the anti-information issue.

Principally, it is a "more speech" solution. Given that bot speech represents a new form of mass communication, adding more information to an already overwhelmed situation is likely to prove ineffective.[109] A solution which is solely a "more speech" solution is not appropriate to addressing the potential for harm bots represented, particularly in their ability to disseminate anti-information.

This option does hold the advantage of being ECHR compliant. In reaching that conclusion, different speech interests relating to bot speech were considered. It was concluded that because of the variety of bot speech, general measures would not be appropriate and restrictions would have to be connected to particular harms. By aligning with bot identity transparency with political speech, a particular potential harm of having elections influenced is addressed.

Despite this advantage, bot identity transparency appears to be inappropriate to address the anti-information issue. If stricter measures were pursued, for example, removal of bots from online platforms, demotion of bot content on social feeds, or a ban on bots, such measures would have to be connected to particular harms, such as electoral and commercial harms addressed by SB-1001 to avoid unjustifiably interfering with other forms of bot speech.[110]

In the context of elections for example, stricter measures could be justified where bots are used for mass communication – particularly if disseminating false information about the electoral process itself. Posts that wrongly inform voters about the electoral process or disparage candidates for office based on fictious information, could be removed and such action could be criminalisation. That appears to be more appropriate at addressing the anti-information than a transparency-improving measure. Particularly given that on the day of an election, there would be insufficient time to respond to the (mass) anti-information bots may spread.

[108] See Chapter 5.1.1.
[109] See Chapter 2.3.3.
[110] See Chapter 5.2.1.

Whilst guaranteeing the prevention of the negative effects of bots in disseminating anti-information may require a technological/industry solution (e.g., for detecting bot speech and responding to it), the State cannot allow private actors to become the arbiters of truth. The role of freedom of expression, for example, in limiting measures against bot speech to particular harms, will evidently be important.

Chapter 6: Microtargeted Advertisements

The role microtargeted advertisements may play in the dissemination of anti-information was demonstrated in Chapter 2. Like with bot speech, such dissemination may overwhelm the idea-consumer, and if made at opportune time, such as before an election, dissemination of anti-information through targeted advertisements could be extremely effective at convincing the idea-consumer of the truth of many ideas.

Firstly, this Chapter will consider the Banning Microtargeted Political Ads Bill. As the name suggests, it is a general measure to prevent the use of microtargeted advertising technology for political purposes. Like with the bot disclosure law in the previous chapter, the compliance with the ECHR of a measure such as banning microtargeted political advertisements will be considered given that there is a lack of proposals addressing microtargeted advertisements in the UK, in order to determine whether it is an option for the UK. It will be concluded that such a measure is Art.10 compliant.

Before considering ECHR compliance however, this Chapter will consider the current UK law that has been applied to microtargeted political advertisements to inform the discussion of whether a ban would be the appropriate approach to protecting against anti-information with regards to freedom of expression.

It will be concluded that banning microtargeted (political) advertisements is the appropriate solution, given that it is a preventative solution, which by focusing on particular harms may avoid a chilling effect on speech whilst also preventing targeted advertisements being used to disseminate timely anti-information which overwhelms the idea-consumer, similarly to how bot speech may be used. Additionally, given that microtargeted advertisements are only available to those with wealth, it will be concluded that a ban is appropriate for equality of arms between speakers in the marketplace of ideas.

Chapter 6.1: Banning Microtargeted Political Ads Bill

Representative Anna Eshoo introduced the Banning Microtargeted Political Ads Bill 6th August 2021, and it was referred to a committee of the House of Representatives that day.[1] It has not progressed any further. The Bill was intended, like the Honest Ads Act, to amend the Federal Election Campaign Act 1971 (FEC). In Section 2(a), the Bill includes a section to be added to the FEC and Section 2(b) extends the definition of "electioneering communication" under the FEC. The final part, 2(c), adds the usual severability clause.

The new section added to the FEC by Section 2(a) would ban 'target[ing] the dissemination of a political advertisement on a covered online platform to an individual, a connected device, or to a group of individuals or connected devices' as well as enabling a third parting to do so.[2] It contains exceptions for targeting by location, to those who give express consent, context advertising (i.e., advertising that is related to information searched for by the user) and random targeting.[3] Additionally, it would create a private right of action for enforcement by individuals,[4] for which injury is automatically constituted by any violation of the ban.[5]

The Bill would affect the "covered online platforms" which are defined as any website, application or advertising network that disseminates political advertisements, excluding however, any such platform that in the 12 months previous 'involved, collected or processed personal information relating to fewer than 50,000,000 individuals,'[6] similarly to the Honest Ads Bill. As such, this potentially suffers from the same issue in that it could miss many microtargeted advertisements. However, if those sites expose

[1] Congress, H.R.4955 - Banning Microtargeted Political Ads Act of 2021 <https://www.congress.gov/bill/117th-congress/house-bill/4955/all-actions> accessed 7 October 2022.
[2] Congress, H.R.4955 - Banning Microtargeted Political Ads Act of 2021 <https://www.congress.gov/bill/117th-congress/house-bill/4955/text> accessed 7 October 2022, see s2(a)(1) "Sec. 325…" (a)(1)-(2).
[3] ibid "Sec. 325" (b)(1)-(4).
[4] ibid "Sec. 325" (c).
[5] ibid "Sec. 325" (c)(1)(C).
[6] ibid "Sec. 325" (d)(3).

users to many more advertisements than other sites, then this will not be a major issue. Research would be required to examine whether this is the case.

"Political advertisements" are defined in the Bill as an 'electioneering communication,' a communication that advocates election or defeat of candidates for Federal office, paid communications that support or attack a candidate for Federal office regardless of whether it advocates voting for or against that candidate, and any advertisement made by or on behalf of a candidate.[7]

Similar First Amendment concerns has been raised of this Bill as with the Honest Ads Bill and SB-1001 the bot disclosure law that suggests that it is not content-neutral as it singles out *political* speech. This will not be addressed in detail here as such concerns could be abated through the same logic employed regarding the content-neutrality of SB-1001 – given that it does not suppress a particular viewpoint, it remains content-neutral.[8]

Chapter 6.2: Microtargeted Ads in the UK

The ICO defines microtargeting as 'a form of online targeted advertising that analyses personal data to identify the interests of a specific audience or individual in order to influence their actions.'[9] This definition was used by the Law Commission and the Scottish Law Commission.[10] The ICO and Electoral Commission have issued sanctions where microtargeted advertisements have been used, however, there is not any law that directly deals with microtargeted advertisements. There is however some existing UK law that has been applied to microtargeted advertisements.

Chapter 6.2.1: Applicable UK Law

Ofcom has a duty to set, review and revise its standards code regarding the content of television and radio programmes.[11] The objectives of the code include upholding the prohibition on political

[7] ibid "Sec. 325" (d)(10)(A)-(D).
[8] See Chapter 5.2.3.
[9] ICO, 'Microtargeting' <https://ico.org.uk/your-data-matters/be-data-aware/social-media-privacy-settings/microtargeting/> accessed 18 October 2022.
[10] Law Commission and Scottish Law Commission, *Electoral Law: A joint final report* (Law Com No. 389, Scot Law Com No 256, 2020) para 12.36.
[11] Communications Act 2003, s319(1).

advertising.[12] Such advertising includes, that which is made by political bodies,[13] that 'which is directed towards a political end,'[14] and that 'which has a connection with an industrial dispute.'[15] Political bodies and political ends includes 'influencing public opinion on a matter which, in the United Kingdom, is a matter of public controversy.'[16]

Beyond Ofcom and the Communications Act 2003, the Information Commissioner's Office (ICO) is responsible for enforcement of a wide range of legislation.[17] Some of the norms contained within the legislation has been used with regards to microtargeted advertisements. Further, some could potentially be used against microtargeted advertising.

In October 2018, the ICO issued a monetary penalty notice for Facebook.[18] Facebook agreed to pay the £500,000 fine in October 2019.[19] The investigation that led to the fine had found that Facebook had failed to protect the personal data of its users,[20] which was harvested by third-party developers,[21] and later used for targeted advertising.[22] Although the fine is relatively small, for a company with a global revenue of £31.5bn the year previous,[23] the ICO was clear that it would have been larger had Facebook's

[12] ibid s319(2)(g).
[13] ibid 321(2)(a).
[14] ibid 321(2)(b).
[15] ibid 321(2)(c).
[16] ibid 321(3)(f).
[17] ICO, 'Legislation we cover' <https://ico.org.uk/about-the-ico/what-we-do/legislation-we-cover/> accessed 16 November 2022.
[18] ICO, 'Investigation into data analytics for political purposes' <https://ico.org.uk/action-weve-taken/investigation-into-data-analytics-for-political-purposes> accessed 16 November 2022.
[19] Alex Hern, 'Facebook agrees to pay fine over Cambridge Analytica scandal' (30 October 2019, The Guardian) <https://www.theguardian.com/technology/2019/oct/30/facebook-agrees-to-pay-fine-over-cambridge-analytica-scandal> accessed 16 November 2022.
[20] ICO, 'Investigation into the use of data analytics in political campaigns: A report to Parliament' (6 November 2018) 26-39 available at <https://ico.org.uk/media/action-weve-taken/2260271/investigation-into-the-use-of-data-analytics-in-political-campaigns-final-20181105.pdf> accessed 16 November 2022.
[21] People that create software to be used with/alongside an online platform.
[22] Jim Waterson, 'UK fines Facebook £500,000 for failing to protect user data' (25 October 2018, The Guardian) <https://www.theguardian.com/technology/2018/oct/25/facebook-fined-uk-privacy-access-user-data-cambridge-analytica> accessed 16 November 2022.
[23] ibid.

"failings" occurred after the General Data Protection Regulation[24] and the Data Protection Act 2018 replaced the Data Protection Act 1998, under which the maximum fine was £500,000.[25]

Chapter 6.2.2: Potential and Proposed Measures

It has been widely suggested and heavily supported that imprinting requirements on "election material" should be extended to cover online advertisements also.[26] Imprinting requirements are currently contained in the Representation of the People Act 1983 (RPA 1983) and the Political Parties, Elections and Referendums Act 2000 (PPERA 2000). The requirements in the RPA 1983 apply to 'any material which can reasonably be regarded as intended to promote or procure the election of a candidate at an election (whether or not it can be so regarded as intended to achieve any other purpose as well,'[27] whilst the requirements of the PPERA 2000 apply to "election material,"[28] which covers the same material as in the RPA but also applies to the material promoting electoral success for political parties, and any candidate or party should they hold or not hold a certain view.[29]

This requirement has already been extended online in Scotland by the Referendums (Scotland) Act 2020 and could be extended by the Secretary of State, who may extend imprinting requirements to "any

[24] European Parliament and Council of the European Union Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/1.
[25] The Data Protection (Monetary Penalties) (Maximum Penalty and Notices) Regulations 2010 SI 2010/31, s2 cf Data Protection Act 1998, s55A(3A)(5) and (9).
[26] Law Commission and Scottish Law Commission (n 9) 12.40.
The Electoral Commission, 'Transparency in digital campaigning: response to Cabinet Office technical consultation on digital imprints' available at <https://www.electoralcommission.org.uk/who-we-are-and-what-we-do/changing-electoral-law/transparent-digital-campaigning/transparency-digital-campaigning-response-cabinet-office-technical-consultation-digital-imprints> accessed 16 November 2022.
Democracy and Digital Technologies Committee, *Digital Technology and the Resurrection of Trust* (HL 2019-21, 77) para 294.
Cabinet Office, *Transparency in digital campaigning: Technical consultation on digital imprints* (2020) 17, available at <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1014665/Digital_imprints_consultation.pdf> accessed 16 November 2022.
Cabinet Office, *Transparency in digital campaigning: Government response* (2021) 10-11, available at <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993905/Digital_imprints_-_FINAL_Government_consultation_response_.pdf> accessed 16 November 2022.
Michael Harker, 'Political advertising revisited: digital campaigning and protecting democratic discourse' (2019) 40 Legal Studies 151-71, 161-2.
[27] RPA 1983, s110(1).
[28] PPERA 2000, s143(1).
[29] ibid s143A(1)(a)-(c).

other material."[30] The Law Commission and Scottish Law Commission recommended that 'The imprint requirement should extend to online campaign material which may reasonably be regarded as intending to procure or promote any particular result.'[31] Additionally, the Government in 2019 went as far as to 'Commit to implementing a digital imprint regime,'[32] but failed to carry out that commitment. Nonetheless, there is significant support for an "Honest Ads" approach. This transparency focused approached is supported also by calls for a searchable database of online political advertisements.[33]

An alternative, may be to restrict,[34] or ban microtargeted advertisements.[35]

At present, some forms of directed marketing are banned in the UK. The Privacy and Electronic Communications (EC Directive) Regulations 2003 (PECR 2003), integrates the similarly named European Community Directive on privacy and electronic communications.[36] The PECR 2003 bans the use of fax machines,[37] unsolicited calls,[38] and email[39] for directed marketing purposes, with exceptions based upon consent.[40]

The ICO has issued fines for contravention of these regulations by groups utilising data that was gathered and shared for both targeted advertisements and email. Leave.EU and GoSkippy, having gathered, shared and utilised personal information for direct marketing by email were fined £75,000

[30] Law Commission and Scottish Law Commission (n 9) 12.40 cf. RPA 1983, s110(2)(b) and (7) and PPERA 2000, s143(6).
[31] ibid 15.77.
[32] Kevin Foster, 'Press release: Government safeguards UK elections' (Gov.uk, 5 May 2019) <https://www.gov.uk/government/news/government-safeguards-uk-elections> accessed 18 October 2022.
[33] Electoral Commission, 'Digital campaigning: Increasing transparency for voters' (June 2018) paras 55-61, available at: <https://www.electoralcommission.org.uk/sites/default/files/pdf_file/Digital-campaigning-improving-transparency-for-voters.pdf> accessed 18 October 2022.
Michela Palese and Josiah Mortimer, 'Reigning in the Political "Wild West": Campaign Rules for the 21st Century (Electoral Reform Society, February 2019) 13, available at: <https://www.electoral-reform.org.uk/latest-news-and-research/publications/reining-in-the-political-wild-west-campaign-rules-for-the-21st-century/> accessed 18 October 2022.
[34] Harker (n 26) 160, for example 'modest interventions (eg stopping lookalike profiling) to a complete prohibition of targeting.'
[35] As recommended by a House of Commons committee – Digital, Culture, Media and Sport Committee, *Disinformation and 'fake news': Interim Report* (HC 2017-19, 363) para 142.
[36] European Parliament and Council of the European Union Directive 2002/58/EC of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) [2002] OJ L201/37.
[37] PECR 2003, SI 2003/2426, reg(20).
[38] ibid reg(21).
[39] ibid reg(22)-(23).
[40] ibid reg(20)(1)(a)-(c) and (2)-(3), reg(21)(1)(a)-(b) and (2)-(5), reg(22)(2), (3)(a)-(c) and (4).

and £60,000 respectively.[41] Should a similar measure be implemented banning microtargeted advertisements except where users consent or consent to their data being used in that manner, the ICO has demonstrated its investigative ability in enforcing such a rule.

Chapter 6.3: Compliance with ECHR of a Microtargeted Advertisement Ban

The proposed measures for tackling microtargeted advertisements in the UK tend to have a transparency-improving approach. Given the criticism that has been laid out in this thesis regarding transparency-improving measures, this section considers the compliance with Art.10 of the suggestion of a ban on microtargeted advertisements.

Given the focus on freedom of expression in this thesis, the analysis will come from an Art.10 perspective, however, given the roll of data-mining in microtargeting, a right to privacy/Art.8, ECHR perspective may be one worthy of consideration elsewhere, particularly given that it would be difficult to conclude expression is protected where it necessitates an interference with the right to privacy. Firstly, whether microtargeted advertisements fall within the scope of Art.10 will be considered before considering whether a general ban on microtargeted advertisements could be justified under Art.10(2). The first part will find that there is reason to question the applicability of Art.10 protection to microtargeted advertisements as an unprecedented form of expression, in either case however, a general ban on microtargeted advertisements could be justified.

Chapter 6.3.1: Art.10 Protection for Microtargeted Ads

Microtargeted advertisements as a form of political expression, prima facie, would appear to be entitled to strong Art.10 protection,[42] particularly if they were made on behalf of a politician.[43] As such it would

[41] ICO, 'Investigation into the use of data analytics in political campaigns: A report to Parliament' (n 20) 66-110.
[42] *TV Vest AS and Rogaland Pensjonistparti v Norway* App. No.21132/05, [64].
[43] Which 'call for the closest scrutiny on the part of the Court' (*Castells v Spain* App. No.11798/85, [42]) where only a narrow margin of appreciation is allowed (*Otegi Mondragon v Spain* App. No.2034/07, [51]).

be reasonable to conclude that 'Paid-for political micro-targeting, as a form of political expression, is therefore a form of political speech under Article 10,'[44] particularly given the 'broad notion of what constitutes an exercise of freedom of expression.'[45]

Whereas other cases considered by the authors in reaching that conclusion concerned criticism directed at political figures or political expression aimed at a public/general audience,[46] the same cannot be said of microtargeting. Microtargeted advertisements more closely resemble proselytising than traditional advertisement or other forms of expression directed at a general audience.

Microtargeted advertisements that are not political also *prima facie* appear to be protected expression, as commercial speech, including advertisements, have been protected previously.[47] Although there is a wide margin of appreciation with regards to the regulation of commercial speech,[48] and such advertisements would not benefit from the heightened level of protection that political speech enjoys.[49]

However, microtargeted advertisements, regardless of content, are an unprecedented form of speech which requires more careful inspection.

The Court has found the internet as providing 'an unprecedent platform for the exercise of freedom of expression,'[50] however it recognises that it brings new dangers also.[51] The Court has subsequently paid particular attention to expression in this new context, for example, it has recognised the citizen journalism impact of YouTube,[52] and applications as a means of imparting information and ideas.[53] The

[44] Tom Dobber, Ronan Ó Fathaigh and Frederik Zuiderveen Borgesius, 'The regulation of online political micro-targeting in Europe' (2019) 8(4) Internet Policy Review <https://policyreview.info/articles/analysis/regulation-online-political-micro-targeting-europe> accessed 1 June 2023.
[45] ibid. Dobber, Fathaigh and Borgesius refer to cases regarding posting on social media (*Einarsson v Iceland* App. No.24703/15, *Mariya Alekhina and Others v Russia* App. No.38004/12), distributing leaflets (*Andrushko v Russia* 4260/04) and displaying posters (*Kandzhov v Bulgaria* App. No.68924/01).
[46] ibid.
[47] See for example, *Sekmadienis Ltd v Lithuania* App. No.69317/14.
[48] See *markt intern Verlag GmbH and Klaus Beermann* App. No.10572/83, [33] and *Casado Coca v Spain* App. No.15450/89, [50] cf *Mouvement Raëlien Suisse v Switzerland* App. No.16354/06, [61], and *Sekmadienis* (n 47) [73].
[49] *Wingrove v UK*, App. No.17419/90, [58].
[50] *Delfi AS v Estonia* App. No.64569/09, [110] cf *Ahmet Yildirim v Turkey* App. No.3111/10, [48] and *Times Newspapers Ltd v UK (Nos 1 and 2)* App. Nos.3002/03 and 23676/03, [27].
[51] *Delfi AS* (n 50) [110].
[52] *Cengiz and Others v Turkey* App. Nos.48226 and 14027/11, [52].
[53] *Magyar Kétfarkú Kutya Part v Hungary* App. No.201/17, [35]-[37].

nature of microtargeted advertisements as expression, due to their particular characteristics of enabling automatic and personalised mass communication, warrant careful inspection. A closer look at microtargeted advertisements would suggest they potentially should not fall within the scope of freedom of expression, as they cannot be equated with traditional forms of advertising and more closely resemble expression like attempts to proselytise individuals.

An individual has a right to attempt to proselytise another individual,[54] but it is limited to an attempt to do so.[55] Repeated attempts to convince an individual to believe in a particular doctrine could amount to harassment.[56] This expression is between individuals, it is not directed to a general/public audience, and it is one-way. Person A speaks to Person B, it is not a discussion or debate.

Microtargeted seek to persuade an individual to vote or purchase as a proselytiser seeks to persuade an individual to believe. They are personalised for and directed at an individual. The expression is essentially private, even though other individuals may receive similar advertisements. Recognising a right to such expression would indicate a right to speak to certain people, or a right to a platform. A legitimate aim would be required to deplatform the advertiser. Regardless, there is no right to a platform.

Expression made in private can be protected by freedom of expression, though expression that the Court has recognised in private often represents strong speech interests, including complaints of public official's conduct being unlawful,[57] complaints about an electoral candidate,[58] and a prison's administration,[59] as well as accusations against a judge.[60] Additionally, it found there to be protected expression where an applicant was criminally convicted for defamation after serving an employee an

[54] *Kokkinakis v Greece* App. No.14307/88, [31].
[55] 'the right to *try* to convince one's neighbour' (emphasis added) *Kokkinakis* (n 54) [31] and *Larissis and Others v Greece* App. No.23372/94, [45].
[56] Equality Act 2010, s26 – subsection 5 lists 'religion or belief' as a protected characteristic for which lack of belief qualifies; 'but it is also a precious asset for atheists, agnostics, sceptics and the unconcerned.' (*Kokkinakis* (n 54) [31]).
[57] *Zakharov v Russia* App. No.14881/03, [8] and [22]-[23], and *Sofranschi v Moldova* App. No.34690/05, [29].
[58] *Sofranschi* (n 57) [29].
[59] *Marin Kostov v Bulgaria* App. No.13801/07, [42].
[60] *Raichinov v Bulgaria* App. No.47579/99 [10] and [43].

official document requesting she perform her legal obligation. [61] Commercial microtargeted advertisements may not have a sufficiently strong speech interest.

Although microtargeted advertisements have a small audience given that they are directed at particular individuals, they are still a form of mass communication that is capable of large impact.

The Court has examined the potential impact of expression in determining the proportionality of interference in cases concerning expression that represents a national security threat.[62]In those cases, the applicants' expression was addressed to a very small audience, and it subsequently had a limited potential to be a national security concern or cause public disorder, and interference was therefore disproportionate.[63] Conversely, microtargeted advertisements may have a limited audience but their potential impact is significant.

Although microtargeted advertisements, particularly political advertisements, appear to fall within the scope of expression which Art.10 protects, there is reason to question whether they would or should be similarly treated by the Court. For the purposes of considering whether a general ban on microtargeted advertisements may be justified, it will be assumed that they may fall within the scope of Art.10.

Chapter 6.3.2: Justification of a General Ban on Microtargeted Advertisements

Article 10 generally applies not only to the content of expression but also the means by which it is expressed/transmitted, since a restriction on the means of expression necessarily interferes with the right to receive and impact information.[64]

In *Animal Defenders International v UK*, that the proportionality of a general measure depends upon, the legislative reasoning for it, the risk of abuse should the general measure be relaxed and whether it

[61] *Matalas v Greece* App. No.1864/18, [9] and [35].
[62] *Karatas v Turkey* App. No.23168/94, and *Polat v Turkey* App. No.23500/94.
[63] ibid [52]-[54] and [47]-[49] respectively.
[64] *Magyar Kétfarkú Kutya Párt v Hungary* App. No.201/17, [87] *Ahmet Yıldırım v Turkey* App. No.3111/10, [50], and *Autronic AG v Switzerland* App. No.12726/87, [47].

is more appropriate for achieving the legitimate aim than a measure that allows for a case-by-case examination.[65]

The legislative reasoning for banning microtargeted ads would be markedly similar to that of the ban on political advertising in broadcasts that was examined in *Animal Defenders*,[66] 'the need to protect the electoral process as part of the democratic order… given the risk posed to the right to free elections.'[67] The State would enjoy a wide margin of appreciation in pursuing this aim.[68] State authorities are in 'a better position than the international judge'[69] to determine what the necessity of banning microtargeted political advertisements would be. Considering the suggestion has been made already,[70] although without a Bill being proposed, it would appear that State authorities have already determined it may be necessary.

Additionally, in *Animal Defenders* the historic context of the measure was relevant in determining the margin of appreciation afforded to the State with regards to the proportionality of the interference with freedom of expression.[71] A ban on microtargeted political advertisements could make use of this historic context. It would be convincing to say that such a ban would represent an updated version of the historic ban, allowing it to continue to be effective in the modern day.

The risk of abuse should the general measure be relaxed would appear to be high given that microtargeted political advertisements have already been employed in the UK, and a general ban would seem to be much more appropriate than an examination on a case-by-case basis by virtue of the fact it would be very burdensome for a public authority to examine the vast array of personalised microtargeted advertisements that internet users in the UK encounter.

Ultimately, such a measure is very likely ECHR compliant.

---

[65] *Animal Defenders International v the United Kingdom* App. No.48876/08, [108].
[66] s321(2) Communications Act 2003, cf. *Animal Defenders* (n 65) [3].
[67] *Animal Defenders* (n 65) [111].
[68] *Bowman v UK* App. No.24839/94, [43].
[69] *Handyside v UK* App.No. 5493/72, [48] cf *Otto-Preminger-Institut v Austria* App. No.13470/87, [56].
[70] Digital, Culture, Media and Sport Committee (n 35).
[71] *Animal Defenders* (n 65).

Chapter 6.4: Appropriateness of a Ban on Microtargeted Advertisements

Banning microtargeted advertisements is an appropriate solution to the anti-information issue. As a preventative measure, which seeks to stop anti-information entering the marketplace of ideas in the first place, it is the preferred approach to the anti-information issue.

The harms of targeted advertisements are known and understood by public authorities in the UK. A ban on microtargeted advertisements addresses those harms and it would be compatible with the ECHR. A ban on microtargeted political advertisements was found to be ECHR compliant however, a ban that addresses other harms identified by the "better placed"[72] public authorities could also be compatible due to the wide margin of appreciation that would be afforded to the State.

A less strict measure, like removal/access limitation or information correction, are likely inappropriate to microtargeted advertisements. Each of those measures would likely dissuade advertisers from using the technology, result in a chilling effect on speech. By limiting the ban to particular harms, less of a chill is applied to speech. No dissuasion of the use of the technology would occur other than harmful misuse, including the dissemination of anti-information.

Additionally, each of those measures would have to be applied carefully to avoid reinforcing belief in anti-information and as with bot speech, the possibility for mass communication that leaves little time for correction to occur is a concern. A ban avoids such concerns, and is therefore the most appropriate solution, in this instance.

Further, a ban on microtargeted advertisements removes advantages held by wealth individuals in the marketplace of ideas. Microtargeted advertisements are only available to wealthy individuals and they represent an inequality of arms between speakers in the marketplace of ideas. A ban would prevent wealth being used to have a disproportionate for on the marketplace of ideas.

[72] *Otto-Preminger-Institut* (n 69).

Chapter 7: Conclusion

Before concluding as to the core argument of this thesis, it is worth highlighting a couple of observations.

Whatever approach to the anti-information issue employs, this thesis demonstrates the need for that approach to be informed by psychological and communicative science research. This research is relevant not only for legislators but also for intermediaries that may have duties imposed upon them to effectively respond to anti-information.

Secondly, free speech narratives with regards to anti-information are often employed against "others," "democratic institutions,"[1] and ultimately, human rights in general. Discussions on how to respond to anti-information may employ the language of "facts" to make assertions contrary to human rights interests. One example the author encountered in researching for this thesis which demonstrates this adequately comes from Baroness Fox who, in debate of the Online Safety Bill, leveraged free speech narratives to make assertions of "fact" about "identity politics," "the use of pronouns," "the biological fact of sex" and "safe space warriors."[2] Similar comments were made by Joanna Cherry in the House of Commons.[3] Those concerned with the impact of anti-information regulation of freedom of expression should be wary of such dangerous narratives and oppose them where possible.

Chapter 7.1: Prevention Strategy

Chapter 2 examined the "truth-seeking" theories, the rationality of the idea-consumer and the role technology plays in the dissemination of false information. This Chapter brought into question whether the truth-seeking theories, the argument from truth and the marketplace of ideas, which justify protecting speech on the grounds that it leads to the discovery of truth, apply to factual truths. It went on to find that whilst the truth-seeking theories often assume the idea-consumer, the person that

[1] See for example, Christopher A Smith, 'Weaponized iconoclasm in Internet memes featuring the expression "Fake News"' (2019) 13(3) Discourse & Communication 303-319.
[2] HL Deb 12 May 2022, vol 822, cols 189-90.
[3] HC Deb 11 May 2022, vol 714, col 189.

encounters ideas and comes to believe them, is a rational actor. A rational actor would help to perform the truth-seeking function, as a rational actor would prefer to believe true ideas over false. However, the idea-consumer was found not to be a rational actor, and instead they were found to have difficulties discerning truth and falsity because of psychological effects that affect the information which the idea-consumer chooses to form their beliefs upon, and that affect how they value information that they encounter. Further, it was also concluded that the idea-consumer passively obtains many of their beliefs and generally does not "test their thinking," Finally, Chapter 2 discussed technologies that exacerbate the psychological effects the idea-consumer faces and in particular, those which leave the idea-consumer, and subsequently the marketplace of ideas, with insufficient time to remedy falsity – bots and targeted advertisements. Ultimately, given that the idea-consumer, by not always favouring the truth as a rational idea-consumer would, is not guaranteed to perform its necessary function in the marketplace of ideas of discerning truth, the assumption that discussion, "more speech" or "counterspeech" can be relied upon to remedy falsity is not applicable to the anti-information issue.

Chapter 3 briefly outlined different solutions that may be applied to the anti-information issue, including "more speech," information correction, removal/access limitation and preventing anti-information being expressed in the first place. It found that prevention was the preferred approach due to the strength of anti-information and the advantages it holds due to new technologies. It emphasised that a "more speech" solution was inapplicable to the anti-information issue and demonstrated how different technologies can be used to take advantage of psychological tendencies of the idea-consumer, to affect their beliefs.

Chapters 4, 5 and 6 considered whether different legislative measures were appropriate solutions to the anti-information issue. It examined the effectiveness of measures discussed in each and made suggestions in light of their different likely effectiveness. The least likely to be effective were "more speech" measures such as bot identity transparency, as discussed in Chapter 5. Whereas the most likely to be effective were preventative measures such as a ban on microtargeted (political) advertisements. The most appropriate solution, that which balanced the aim of protecting against the harms of anti-information and relevant speech interests, was preventative approaches to the anti-information issue as

suggested in the conclusions of Chapters 4, 5 and 6. A solution which prevents anti-information entering the marketplace of ideas in the first place is necessary to avoid anti-information's harms, particularly because new technologies, like bots and targeted advertisements, allow anti-information to be more convincing than ever before by overwhelming the idea-consumer with anti-information, including that which is personalised to them.

Bibliography

Legislation and Bills

United Kingdom

- Communications Act 2003.
- Data Protection Act 1998.
- Elections Act 2022.
- Fraud Act 2006.
- Human Rights Act 1998.
- Online Safety Bill HL Bill (2022-23) 87(Rev).
- Online Safety HC Bill (2022-23) [209].
- Political Parties, Elections and Referendums Act 2000.
- Representation of the People Act 1983.
- The Data Protection (Monetary Penalties) (Maximum Penalty and Notices) Regulations 2010 SI 2010/31.
- The Elections Act 2022 (Commencement No. 8) Regulations 2023, SI 2023/552.
- The Elections Act 2022 (Commencement No.1 and Saving Provision) Regulations 2022, SI 2022/908.
- The Privacy and Electronic Communications (EC Directive) Regulations 2003, SI 2003/2426.

United States of America (Federal)

- Congress, H.R.2592 – Honest Ads Act, <https://www.congress.gov/bill/116th-congress/house-bill/2592/ > accessed 15 January 2022.
- Congress, H.R.4955 - Banning Microtargeted Political Ads Act of 2021 <https://www.congress.gov/bill/117th-congress/house-bill/4955/all-actions> accessed 7 October 2022.

- Congress, S.1356 - A bill to enhance transparency and accountability for online political advertisements by requiring those who purchase and publish such ads to disclose information about the advertisements to the public, and for other purposes, <https://www.congress.gov/bill/116th-congress/senate-bill/1356/ > accessed 15 January 2022.
- Congress, S.1989 – Honest Ads Act, <https://www.congress.gov/bill/115th-congress/senate-bill/1989/ > accessed 15 January 2022.
- Federal Election Campaign Act Amendments of 1974, Pub L 93-433.

United States of America (California)

- Business and Professions Code (California).
- SB-1001 <https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1001> accessed 1 January 2022.

Other

- European Parliament and Council of the European Union Directive 2002/58/EC of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) [2002] OJ L201/37 (European Union).
- European Parliament and Council of the European Union Regulation (EU) 2016/679 of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/1. (European Union).
- LOI n° 2018-1202 du 22 décembre 2018 relative à la lutte contre la manipulation de l'information (1) available at <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000037847559> accessed 24 April 2023 (France).

Articles

- Asbjørn Følstad, Cecilie Bertinussen Nordheim & Cato Alexander Bjørkli, 'What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study' (2018) Internet Science, available at <https://link.springer.com/chapter/10.1007/978-3-030-01437-7_16> accessed 27 April 2023.
- Abby Wood and Ann Ravel, 'Fool Me Once: Regulating "Fake News" and Other Online Advertising' (2018) 91(6) Southern California Law Review 1223-1278.
- Ajnesh Prasad, 'Denying Anthropogenic Climate Change: Or, How Our Rejection of Objective Reality Gave Intellectual Legitimacy to Fake News' (2019) 34(SI) Sociological Forum 1217-1234.
- Alberto Ardèvol-Abreu and Homero Gil de Zúñiga, 'Effects of Editorial Media Bias Perception and Media Trust on the Use of Traditional, Citizen, and Social Media News' (2017) 94(3) Journalism & Mass Communication Quarterly 703-724.
- Alessandro Bessi and Emilio Ferrara, 'Social bots distort the 2016 U.S. Presidential election online discussion' (2016) 21(11) First Monday <https://firstmonday.org/ojs/index.php/fm/article/view/7090> accessed 22 April 2023.
- Alexandra Maftei, Andrei-Corneliu Holman, Ioan-Alex Merlici, 'Using fake news as means of cyber-bullying The link with compulsive internet use and online more disengagement' (2022) 127 Computers in Human Behaviour <https://www.sciencedirect.com/science/article/abs/pii/S0747563221003551> accessed 13 June 2023.
- Alison R. Fragale and Chip Heath, 'Evolving Informational Credentials: The (Mis)Attribution of Believable Facts to Credible Sources' (2004) 30(2) Personality and Social Psychology Bulletin 225-236.
- Amartya Sen, 'Rational Fools: A Critique of the Behavioral Foundations of Economic Theory' (1977) 6(4) Philosophy & Public Affairs 317-344.

- Amon Rapp, Lorenzo Curti and Arianna Boldi, ‘The human side of human-chatbot interaction: A systemic literature review of ten years of research on text-based chatbots’ (2021) 151 International Journal of Human-Computer Studies, available at <https://www.sciencedirect.com/science/article/pii/S1071581921000483> accessed 27 April 2023.
- Amos Tversky and Daniel Kahneman, ‘The Framing of Decisions and the Psychology of Choice’ (1981) 211(4481) Science 453-458.
- Andras Koltay, ‘Constitutional protection of lies?’ (2020) 25(3) Communications Law 131-149.
- Andrew Moshirnia, ‘Who Will Check the Checkers? False Factcheckers and Memetic Misinformation’ (2020) 4 Utah Law Review 1029-74.
- Anthony G. Greenwald & Mahzarin R. Banaji, ‘Implicit social cognition: Attitudes, self-esteem, and stereotypes’ (1995) 102(1) Psychological Review 4-27.
- Anthony G. Greenwald and Lina Hamilton Krieger, ‘Implicit Bias: Scientific Foundations’ (2006) 94(4) California Law Review 945-967.
- Anya Schiffrin, ‘Disinformation and Democracy: The Internet Transformed Protest but Did Not Improve Democracy’ (2017) 71(1) Journal of International Affairs 117-126.
- Ari Ezra Waldman, ‘The Marketplace of Fake News’ (2018) 20 University of Pennsylvania Journal of Constitutional Law 845-70.
- Ariella Kristal and Laurie Santos, ‘G.I. Joe Phenomena: Understanding the Limits of Metacognitive Awareness on Debiasing’ (2021) Harvard Business School Working Paper 21-084, 3 <https://www.hbs.edu/faculty/Pages/item.aspx?num=59722> accessed 14 June 2023.
- Barbara Luppi & Francesco Parisi, ‘Beyond Liability: Correcting Optimism Bias through Tort Law’ (2009) 35(1) Queen’s Law Journal 47-66.
- Barry Stricke, ‘People v. Robots: A Roadmap for Enforcing California's New Online Bot Disclosure Act’ (2020) 22 Vanderbilt Journal of Entertainment and Technology Law 839.

- Binxuan Huang and Kathleen M Carley, 'Disinformation and Misinformation on Twitter during the Novel Coronavirus Outbreak' <https://arxiv.org/pdf/2006.04278.pdf> accessed 14 April 2023.
- Brendan Nyhan and Jason Reifler, 'When Corrections Fail: The Persistence of Political Misperceptions' (2010) 32(2) Political Behaviour 303-330.
- Brian A. Nosek, Frederick L. Smyth, Jeffrey J. Hansen, Thierry Devos, Nicole M. Lindner, Kate A. Ranganath, Colin Tucker Smith, Kristina R. Olson, Dolly Chugh, Anthony G. Greenwald and Mahzarin R. Banaji, 'Pervasiveness and correlates of implicit attitudes and stereotypes' (2007) 18 European Review of Social Psychology 36-88.
- Brian Beyersdorf, 'Regulating the "Most Accessible Marketplace of Ideas in History": Disclosure Requirements in Online Political Advertisements After the 2016 Election'(2019) 107(3) California Law Review 1061-1100.
- Cameron Bunker and Michael Varnum, 'How strong is the association between social media use and false consensus?' (2021) 125 Computers in Human Behavior <https://www.sciencedirect.com/science/article/pii/S0747563221002703> accessed 22 April 2023.
- Chadly Stern, Tessa West and Peter Schmitt, 'The Liberal Illusion of Uniqueness' (2014) 25(1) Psychological Science 137-144.
- Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini & Filippo Menczer, 'The spread of low-credibility content by social bots' (2018) 9 Nature Communications 5 <https://www.nature.com/articles/s41467-018-06930-7> accessed 22 April 2023.
- Christopher A Smith, 'Weaponized iconoclasm in Internet memes featuring the expression "Fake News"' (2019) 13(3) Discourse & Communication 303-319.
- Christopher Wonnell, 'Truth and the Marketplace of Ideas' (1986) 19(3) UC Davis Law Review 669-728.

- Daniel Bar-Tal, Alona Raviv and Tali Freund, ‘An Anatomy of Political Beliefs: A Study of Their Centrality, Confidence, Contents, and Epistemic Authority’ (1994) 24(10) Journal of Applied Social Psychology 849-872.
- Daniel Kahneman, Jack L. Knetsch, and Richard H. Thaler, ‘The Endowment Effect, Loss Aversion and Status Quo Bias’ (1991) 5(1) Journal of Economic Perspectives 193-206.
- Daniel T. Gilbert, ‘How Mental Systems Believe‘ (1991) 46(2) American Psychologist 107-119.
- Daniela Manzi, ‘Managing the Misinformation Marketplace: The First Amendment and the Fight Against Fake News’ (2019) 87(6) Fordham Law Review 2623-2651.
- Danielle C. Polage, ‘Making up History: False Memories of Fake News Stories’ (2012) 8(2) Europe’s Journal of Psychology 245-250.
- Dan-Olof Rooth, ‘Implicit Discrimination in Hiring: Real World Evidence’ (2007) IZA Discussion Paper 2764.
- Darren Bush, ‘The Marketplace of Ideas: Is Judge Posner Chasing Don Quixote’s Windmills?’ (2000) 32(4) Arizona State Law Journal 1107-48.
- David A. Broniatowski, Amelia M. Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C. Quinn and Mark Dredze, ‘Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate’ (2018) 108(10) American Journal of Public Health 1378-1384.
- Dawn Carla Nunziato, ‘The Marketplace of Ideas Online’ (2019) 94 Notre Dame Law Review 1519.
- Derek Bambauer, ‘Shopping Badly: Cognitive Biases, Communications, and the Fallacy of the Marketplace of Ideas’ (2006) 77(3) University of Colorado Law Review 649-710.
- Ditte Marie Munch-Jurisic, ‘The Right to Feel Comfortable: Implicit Bias and the Moral Potential of Discomfort’ (2020) 23(1) Ethical Theory and Moral Practice 237-250.
- DR Harris, ‘The development of socio-legal studies in the United Kingdom’ (1983) 3(3) Legal Studies 315-33.

- Edda Humprecht, ‘Where “fake news” flourishes: a comparison across four Western democracies’ (2019) 22(13) Information, Communication & Society 1973-88.
- Edda Humprecht, Frank Esser, and Peter Van Aelst, ‘Resilience to Online Disinformation: A Framework for Cross-National Comparative Research’ (2020) 25(3) The International Journal of Press/Politics 493-516.
- Eiríkur Bergmann, ‘Populism and the politics of misinformation’ (2020) 21(3) The Journal of South African and American Studies 251-65.
- Elena Druică, Fabio Musso and Rodica Ianole-Călin, ‘Optimism Bias during the Covid-19 Pandemic: Empirical Evidence from Romania and Italy’ (2020) 11(3) Games 1 <https://www.mdpi.com/2073-4336/11/3/39> accessed 22 April 2023.
- Ellen Goodman and Lyndsey Wajert, ‘The Honest Ads Act Won’t End Social Media Disinformation, But It’s A Start’ (unpublished) <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3064451> accessed 26 August 2022.
- Emilio Ferrara, ‘Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election’ (2017) 22(8) First Monday <https://firstmonday.org/ojs/index.php/fm/article/view/8005> accessed 22 April 2023.
- Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini (2016) 59(7) Communications of the ACM 96-104.
- Emily Pronin, Daniel Lin, Lee Ross, ‘The Bias Blind Spot: Perceptions of Bias in Self Versus Others’ (2002) 28(3) Personality and Social Psychology Bulletin 369-381.
- Erik Cambria, Praphul Chandra, Avinash Sharma, and Amir Hussain, ‘Do Not Feel The Trolls’ (2010) 664 CEUR Workshop Proceedings 1.
- Erin Buckels, Paul Trapnell, and Delroy Paulhus, ‘Trolls just want to have fun’ (2014) 67 Personality and Individual Differences 97-102.
- Farhana Sultana, ‘The false equivalence of academic freedom and free speech: Defending academic integrity in the age of white supremacy, colonial nostalgia, and anti-intellectualism’ (2018) 17(2) ACME: An International Journal for Critical Geographies 228-57.

- Frederick Schauer, ‘Facts and the First Amendment’ (2010) 57 UCLA Law Review 897-919.
- Frederick Schauer, ‘The Second-Best First Amendment’ (1989) 31(1) William & Mary Law Review 1-23.
- Geoffrey Beattie, ‘Optimism bias and climate change’ (2018) 33 British Academy Review 12-15.
- Gordon Pennycook, Tyrone D. Cannon and David G. Rand, ‘Prior exposure increases perceived accuracy of fake news’ (2018) 147(12) Journal of Experimental Psychology 1865-80.
- Gregory Brazeal, ‘How much does a belief cost?: Revisiting the Marketplace of Ideas’ (2011) 21(1) Southern California Interdisciplinary Law Journal 1-46.
- Heather Ford, Elizabeth Dubois, and Cornelius Puschmann, ‘Keeping Ottawa Honest–One Tweet at a Time? Politicians, Journalists, Wikipedians and Their Twitter Bots’ (2016) 10 International Journal of Communication 4891-4914.
- Jarred Prier, ‘Commanding the Trend: Social Media as Information Warfare’ (2017) 11(4) Strategic Studies Quarterly 50-85.
- Jay Hmielowski, Sarah Staggs, Myiah Hutchens, and Michael Beam, ‘Talking Politics: The Relationship Between Supportive and Opposing Discussion With Partisan Media Credibility and Use’ (2022) 49(2) Communications Research 221-244.
- Jennifer M Grygiel, ‘Algorithmic propaganda: how Facebook meddles with democracy’ (2020) 25(1) Communications Law 23-30.
- Jens Asendorpf, Rainer Banse, and Daniel Mücke, ‘Double Dissociation Between Implicit and Explicit Personality Self-Concept: The Case of Shy Behavior’ (2002) 83(2) Journal of Personality and Social Psychology 380-393.
- Jerry Kang, Mark Bennett, Devon Carbado, Pam Casey, Nilanjana Dasgupta, David Faigman, Rachel Godsil, Anthony G. Greenwald, Justin Levinson, Jennifer Mnookin, ‘Implicit Bias in the Courtroom’ (2012) 59(5) UCLA Law Review 1124-1886.

- Jessica Feezell, Meredith Conroy, Barbara Gomez-Aguinaga and John Wagner, ‘Who Gets Flagged? An Experiment On Censorship and Bias in Social Media Reporting’ (2023) 56(2) Political Science & Politics 222-26.
- Joan Costa-Font, Elias Mossialos, and Caroline Rudisill, ‘Optimism and the perceptions of new risks’ (2009) 12(1) Journal of Risk Research 27-41.
- John Dovidio, Kerry Kawakami and Samuel Gaertner, ‘Implicit and Explicit Prejudice and Interracial Interaction’ (2002) 82(1) Journal of Personality and Social Psychology 62-68.
- John Frank Weaver, ‘Everything Is Not Terminator: We Need the California Bot Bill But We Need It to Be Better’ (2018) 1(6) The Journal of Robotics, Artificial Intelligence & Law 431.
- John King, ‘Microtargeted Political Ads: An Intractable Problem’(2022) 102(3) Boston University Law Review 1129-67.
- John-Stewart Gordon and Ausrine Pasvenkiene, ‘Human rights for robots? A literature review’ (2021) 1(4) AI and Ethics 579-591.
- Karla Dhungana Sainju, Huda Zaidi, Niti Mishra and Akosua Kuffour, ‘Xenophobic Bullying and COVID-19: An Exploration Using Big Data and Qualitative Analysis’ (2022) 19(8) International Journal of Environmental Research and Public Health 4824.
- Kevin Corti and Alex Gillespie, ‘Co-constructing intersubjectivity with artificial conversational agents: People are more likely to initiate repairs of misunderstandings with agents represented as human’ (2016) 58 Computers in Human Behavior 431-442.
- Kien Hoa Ly, Ann-Marie Ly and Gerhard Andersson, ‘A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods’ (2017) 10 Internet Interventions 39-46.
- Kimberly A Wade-Benzoni, Ann E Tenbrunsel and Max H Bazerman, ‘Egocentric Interpretations of Fairness in Asymmetric, Environmental Social Dilemmas: Explaining Harvesting Behavior and the Role of Communication’ (1996) 67(2) Organizational Behavior and Human Decision Processes 111-26.

- Lars Thøger Christensen and George Cheney, ‘Peering into Transparency: Challenging Ideals, Proxies, and Organizational Practices’ (2015) 25(1) Communication Theory 70-90.
- Laurie A. Rudman, ‘Sources of Implicit Attitudes’ (2004) 13(2) Current Directions in Psychological Science 79-82.
- Lee Ross and Constance Stillinger, ‘Barriers to Conflict Resolution’ (1991) 7(4) Negotiation Journal 389-404.
- Lee Ross, David Greene and Pamela House, ‘The “false consensus effect”: An egocentric bias in social perception and attribution processes’ (1977) 13(3) Journal of Experimental Social Psychology 279-301.
- Leslie Gielow Jacobs, ‘Freedom of Speech and Regulation of Fake News’ (2022) 70(S1) The American Journal of Comparative Law i278-i311.
- Linda Babcock and George Loewenstein, ‘Explaining Bargaining Impasse: The Role of Self-Serving Biases’ (1997) 11(1) Journal of Economic Perspectives 109-26.
- Lisa K. Fazio, Nadia M. Brashier, B. Keith Payne and Elizabeth J. Marsh, ‘Knowledge Does Not Protect Against Illusory Truth’ (2015) 144(5) Journal of Experimental Psychology 993-1002.
- Luigi Castelli, Cristina Zogmaister and Silvia Tomelleri, ‘The Transmission of Racial Attitudes Within the Family’ (2009) 45(2) Developmental Psychology 586-91.
- Lynn Hasher, David Goldstein and Thomas Toppino, ‘Frequency and the Conference of Referential Validity’ (1977) 16(1) Journal of Verbal Learning and Verbal Behavior 107-112.
- Madeline Lamo & Ryan Calo, ‘Regulating Bot Speech’ (2019) 66(4) UCLA Law Review 988-1028.
- Magdalena Saldaña, Shannon C McGregor, and Homero Gil De Zúñiga, ‘Social Media as a Public Space for Politics: Cross-National Comparison of News Consumption and Participatory Behaviours in the United States and the United Kingdom’ (2015) 9 International Journal of Communication 3304-3326.

- Magdalena Wojcieszak, ‘Computer-Mediated False Consensus: Radical Online Groups, Social Networks and News Media’(2011) 14(4) Mass Communication and Society 527-546.
- Magdalena Wojcieszak, ‘False Consensus Goes Online: Impact of Ideologically Homogeneous Groups on False Consensus’ (2008) 72(4) Public Opinion Quarterly 781-91.
- Mahzarin R. Banaji and Anthony G. Greenwald, ‘Blindspot: Hidden Biases of Good People’ (2013 Delacorte Press).
- Marco T. Bastos and Dan Mercea, ‘The Brexit Botnet and User-Generated Hyperpartisan News’ (2017) 37(1) Social Science Computer Review 38-54.
- Marina Azzimonti and Marcos Fernandes, ‘Social media networks, fake news, and polarization’ (2023) 76 European Journal of Political Economy 23 <https://www.sciencedirect.com/science/article/pii/S0176268022000623> accessed 22 April 2023.
- Marita Skjuve, Ida Maria Haugstveit, Asbjørn Følstad and Petter Bae Brandtzaeg, ‘Help! Is my chatbot falling into the uncanny valley? An empirical study of user experience in human–chatbot interaction’ (2019) 15(1) Human Technology 30-54.
- Marni Soupcoff, ‘Honest Ads in the Agora’ (2017) 40 Regulation 80.
- Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi and Michele Starnini, ‘The echo chamber effect on social media’ (2021) 118(9) Proceedings of the National Academy of Sciences of the United States of America 1-8.
- Matthew Davis and Per Fors, ‘Towards a Typology of Intentionally Inaccurate Representations of Reality in Media Content’ (2020) 590 IFIP Advances in Information and Communication Technology 291-304.
- Matthew Hines, ‘I Smell a Bot: California's SB 1001, Free Speech, and the Future of Bot Regulation’ (2019) 57 Houston Law Review 405.
- Matthew Motta, ‘The Dynamics and Political Implications of Anti-Intellectualism in the United States’ (2018) 46(3) American Politics Research 465-498.

- Michael Cacciatore, ‘Misinformation and public opinion of science and health: Approaches, findings and future directions’ (2021) 118(15) Proceedings of the National Academy of Sciences of the United States of America 1-8.
- Michael Harker, ‘Political advertising revisited: digital campaigning and protecting democratic discourse’ (2019) 40 Legal Studies 151-71.
- Michael Johann and Lars Bülow, ‘One Does Not Simply Create a Meme: Conditions for the Diffusion of Internet Memes’ (2019) 13 International Journal of Communication 1720-1742.
- Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H Eugene Stanley, Walter Quattrociochi, ‘The spreading of misinformation online’ (2016) 113(3) Proceedings of the National Academy of Sciences of the United States of America 554-559.
- Mika Westerlund, ‘The Emergence of Deepfake Technology: A Review’ (2019) 9(11) Technology Innovation Management Review 40-53.
- Mitchell Rabinowitz, Lauren Latella, Chadly Stern and John Jost, ‘Beliefs about Childhood Vaccination in the United States: Political Ideology, False Consensus, and the Illusion of Uniqueness’ (2016) 11(7) PLoS ONE <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0158382> accessed 13 June 2023.
- Moti Nissani and Donna Marie Hoefler-Nissani, ‘Experimental Studies of Belief Dependence of Observations and of Resistance to Conceptual Change’ (1992) 9(2) Cognition and Instruction 97-111.
- Nicoleta Corbu, Denisa-Adriana Oprea, Elena Negrea-Busuioc and Loredana Radu, ‘“They can’t fool me, but they can fool the others!” Third person effect and fake news detection’ (2020) 35(2) European Journal of Communication 165-180.
- Nicollas de Oliveira, Pedro Pisa, Martin Andreoni Lopez, Dianne Scherly de Medeiros and Diogo Mattos, ‘Identifying Fake News on Social Networks Based on Natural Language Processing: Trends and Challenges’ (2021) 12(1) Information 38.

- Nikos Antonopoulos, Andreas Veglis, Antonis Gardikiotis, Rigas Kotsakis, and George Kalliris, 'Web Third-person effect in structural aspects of the information on media websites' (2015) 44 Computers in Human Behavior 48-58.
- Nilanjana Dasgupta and Anthony G. Greenwald, 'On the Malleability of Automatic Attitudes: Combating Automatic Prejudice with Images of Admired and Disliked Individuals' (2001) 81(5) Journal of Personality and Social Psychology 800-14.
- Nilanjana Dasgupta, 'Implicit Ingroup Favoritism, Outgroup Favoritism, and Their Behavioral Manifestations' (2004) 17(2) Social Justice Research 143-69.
- Patricia G. Devine, 'Stereotypes and Prejudice: Their Automatic and Controlled Components' (1989) 56(1) Journal of Personality and Social Psychology 5-18.
- Paul Brietzke, 'How and Why the Marketplace of Ideas Fails' (1997) 31(3) Valparaiso University Law Review 951-969.
- Paul H. Brietzke, 'Urban Development and Human Development' (1992) 25(3) Indiana Law Review 741-98.
- Paul Wragg, 'Tackling online harms: what good is regulation?' (2019) 2 Communications Law 49-51.
- Peter Fernandez, 'The technology behind fake news' (2017) 34(7) Library Hi Tech News 1-5.
- Phillip Napoli, 'What If More Speech Is No Longer the Solution? First Amendment Theory Meets Fake News and the Filter Bubble' (2018) 70 Federal Communications Law Journal 55-104.

- Phillip Napoli, 'What If More Speech Is No Longer the Solution? First Amendment Theory Meets Fake News and the Filter Bubble' (2018) 70 Federal Communications Law Journal 55-104.

- Pichaya Winichakul, 'The Missing Structural Debate: Reforming Disclosure of Online Political Communications' (2018) 93(5) New York University Law Review 1387.
- PR Chamberlain, 'Twitter as a Vector for Disinformation' (2010) 9(1) Journal of Information Warfare 11-17.

- Pratiwi Utami, ‘Hoax in Modern Politics: The Meaning of Hoax in Indonesian Politics and Democracy’ (2018) 22(2) Jurnal Ilmu Sosial Dan Ilmu Politik 85-97.
- R. J. Dolan, ‘Emotion, Cognition and Behaviour’ (2002) 298 Science 1191-4.
- Rachael Craufurd Smith, ‘Fake news, French Law and democratic legitimacy: lessons for the United Kingdom?’ (2019) 11(1) Journal of Media Law 52-81.
- Ramón Salaverría, Nataly Buslón, Fernando López-Pan, Bienvenido León, Ignacio López-Goñi, and María-Carmen Erviti, ‘Desinformación en tiempos de pandemia: tipología de los bulos sobre la Covid-19’ (2020) 29(3) Profesional de la información 1-17.
- Raúl Rodríguez-Ferrándiz, Cande Sánchez-Olmos, Tatiana Hidalgo-Marí and Estela Saquete-Boro, ‘Memetics of Deception: Spreading Local Meme Hoaxes during COVID-19 1st Year’ (2021) 13(6) Future Internet <https://www.mdpi.com/1999-5903/13/6/152> accessed 22 April 2023.
- Raymond Boudon, ‘Limitations of Rational Choice Theory’ (1998) 104(3) American Journal of Sociology 817-28.
- Raymond Nickerson, ‘Confirmation Bias: A Ubiquitous Phenomenon in Many Guises’ (1998) 2(2) Review of General Psychology 175-220.
- Rebecca Helm and Hitoshi Nasu, ‘Regulatory Responses to ‘Fake News’ and Freedom of Expression: Normative and Empirical Evaluation’ (2021) 21(2) Human Rights Law Review 302-2.
- Richard Posner, ‘Rational Choice, Behavioural Economics, and the Law’ (1997) 50 Stanford Law Review 1551.
- Richard Spearman, ‘Fake news and financial market blues’ (2017) 8 Journal of International Banking and Financial Law 488-90.
- Robert Gorwa and Douglas Guilbeault, ‘Unpacking the Social Media Bot: A Typology to Guide Research and Policy’ (2020) 12(2) Policy & Internet 225-248.
- Robert Luzsa & Susanne Mayr, ‘False Consensus in the Echo Chamber: Exposure to Favorably Biased Social Media News Feeds Leads to Increased Perception of Public Support

for Own Opinions' (2021) 15(1) Cyberpsychology: Journal of Psychosocial Research on Cyberspace <https://cyberpsychology.eu/article/view/12254> accessed 22 April 2023.

- Robert Post, 'The Constitutional Concept of Public Discourse: Outrageous Opinion, Democratic Deliberation, and Hustler Magazine v. Falwell' (1990) 103(3) Harvard Law Review 601-86.
- Robert Vallone, Lee Ross and Mark Lepper, 'The Hostile Media Phenomenon: Biased Perception and Perceptions of Media Bias in Coverage of the Beirut Massacre' (1985) 49(3) Journal of Personality and Social Psychology 577-585.
- Roberta De Cicco, Susanna Cristina Lima da Costa e Silva and Riccardo Palumbo, 'Should a Chatbot Disclose Itself? Implications for an Online Conversational Retailer' (2021) Conversations 2020, available at <https://link.springer.com/chapter/10.1007/978-3-030-68288-0_1> accessed 27 April 2023.
- S Mo Jang and Joon K Kim, 'Third person effects of fake news: Fake news regulation and media literacy interventions' (2018) 80 Computers in Human Behavior 295-302.
- S Mo Jang, Brooke W Mckeever, Robert Mckeever, and Joon Kyoung Kim, 'From Social Media to Mainstream News: The Information Flow of the Vaccine-Autism Controversy in the US, Canada and the UK' (2019) 34(1) Health Communication 110-117.
- Sabina Schnell, 'Transparency in a "Post-Fact" World' (2022) 5(3) Perspectives on Public Management and Governance 222-231.
- Sabine Pahl, Stephen Sheppard, Christine Boomsma, and Christopher Groves, 'Perceptions of time in relation to climate change' (2014) 5(3) WIREs Climate Change 375-88.
- Samantha Bradshaw and Philip N Howard, 'The Global Organization of Social Media Disinformation Campaigns' (2018) 71(1.5) Journal of International Affairs 23-32.
- Samuel C. Woolley and Philip N. Howard, 'Political Communication, Computation Propaganda, and Autonomous Agents' (2016) 10 International Journal of Communication 4882-4890.

- Samuel Rhodes, ‘Filter Bubbles, Echo Chambers, and Fake News: How Social Media Conditions Individuals to Be Less Critical of Political Misinformation’ (2022) 39(1) Political Communication 1-22.
- Sang Ah Kim, ‘Social Media Algorithms: Why You See What You See’ (2017) 2(1) Georgetown Law Technology Review 147-154.
- Sarah Myers West, ‘Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms’ (2018) 20(11) New Media & Society 4366-4383.
- Savvas Zannettou, Tristan Caulfield, William Setzer, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn, ‘Who Let The Trolls Out? Towards Understanding State-Sponsored Trolls’ (2019) WebSci '19: Proceedings of the 10th ACM Conference on Web Science 353-362.
- Shuo Tang, Lars Willnat and Hongzhong Zhang, ‘Fake news, information overload and the third-person effect in China’ (2021) 6(4) Global Media and China 492-507.
- Simon Gächter and Elke Renner, ‘Leaders as role models and “belief managers” in social dilemmas’ (2018) 154 Journal of Economic Behavior & Organization 321-334.
- Soroush Vosoughi, Deb Roy, and Sinan Aral, ‘The Spread of True and False News Online’ (2018) Science 1146.
- Stefan Stieglitz, Florian Brachten, Björn Ross, and Anna Jung, ‘Do Social Bots Dream of Electric Sheep? A Categorisation of Social Media Bot Accounts’ (2017) ACIS 2017 Proceedings 89.
- Stefania Milan, ‘When Algorithms Shape Collective Action: Social Media and the Dynamics of Cloud Protesting’ (2015) 1(2) Social Media + Society 1-10.
- Stephan Lewandowsky & Sander van der Linden, ‘Counter Misinformation and Fake News Through Inoculation and Prebunking’ (2021) 32(2) European Review of Social Psychology 348-384.
- Stephen J Shapiro, ‘Comparing Free Speech: United States v. United Kingdom’ (1989) 19(2) University of Baltimore Law Forum 17-20.

- Sue Curry Jansen and Brian Martin, ‘The Streisand Effect and Censorship Backfire’ (2016) 9 International Journal of Communication 656-71.
- Susan Goldstein, Noni E MacDonald, Sherine Guirguis, ‘Health communication and vaccine hesitancy’ (2015) 33(34) Vaccine 4212-4214.
- Taavi Annus, ‘Comparative Constitutional Reasoning: The Law and Strategy of Selecting the Right Arguments’ (2004) 14(2) Duke Journal of Comparative & International Law 301-50.
- Tali Sharot, ‘The optimism bias’ (2011) 21(23) Current Biology R941-R945.
- Tawanna Lee, ‘Combating Fake News with “Reasonable Standards”’ (2021) 43(1) Hastings Communications and Entertainment Law Journal 81-107.
- Ted de Boer, 'Vergelijkenderwijs: de inspiratie van buitenlands recht' (1992) 123(6033) Weekblad Voor Privaatrecht Notariaat en Registratie 39-48.
- Tetyana Lokot and Nicholas Diakopoulous, ‘News Bots: Automating news and information dissemination on Twitter’ (2016) 4(6) Digital Journalism 682-699.
- Thomas Scanlon, ‘A Theory of Freedom of Expression’ (1972) 1(2) Philosophy & Public Affairs 204-226.
- Tim Hwang, Ian Pearce, and Max Nanis, ‘Socialbots: Voices from the Fronts’ 19(2) Interactions 38-45.
- Tim Wood and Melissa Aronczyk, ‘Publicity and Transparency’ (2020) 64(11) American Behavioral Scientist 1531-1544.
- Timo Harjuniemi, ‘Post-truth, fake news and the liberal “regime of truth” – The double movement between Lippmann and Hayek’ (2022) 37(3) European Journal of Communication 269-283.
- Tom Dobber, Ronan Ó Fathaigh and Frederik Zuiderveen Borgesius, ‘The regulation of online political micro-targeting in Europe’ (2019) 8(4) Internet Policy Review <https://policyreview.info/articles/analysis/regulation-online-political-micro-targeting-europe> accessed 1 June 2023.

- Ulises Meijas and Nikolai Vokuev, ‘Disinformation and the media: the case of Russia and Ukraine’ (2017) 39(7) Media, Culture & Society 1027-42.
- Valentina Vellani, Sarah Zheng, Dilay Ercelik, Tali Sharot, ‘The illusory truth effect leads to the spread of misinformation’ (2023) 236 Cognition 6, <https://www.sciencedirect.com/science/article/pii/S0010027723000550> accessed 22 April 2023.
- Vivian Ta, Caroline Griffith, Carolynn Boatfield, Xinyu Wang, Maria Civitello, Haley Bader, Esther DeCero and Alexia Loggarakis, ‘User Experiences of Social Support From Companion Chatbots in Everyday Contexts:: Thematic Analysis’ (2020) 22(3) Journal of Medical Internet Research, available at <https://www.jmir.org/2020/3/e16235/> accessed 27 April 2023.
- W Lance Bennett and Steven Livingston, ‘The disinformation order: Disruptive communication and the decline of democratic institutions’ (2018) 33(2) European Journal of Communication 122-139.
- W Phillips Davison, ‘The Third-Person Effect in Communication’ (1983) 47(1) Public Opinion Quarterly 1-15.
- Weiyan Shi, Xeuwei Wang, Yoo Jung Oh, Jingwen Zhang, Saurav Sahay and Zhou Yu, ‘Effects of Persuasive Dialogues: Testing Bot Identities and Inquiry Strategies’ (2020) Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems 1-13.
- Wilhelm Hofmann and Christopher J. Wilbur, ‘Are “implicit” attitudes unconscious?’ (2006) 15(3) Consciousness and Cognition 485-99.
- William Gehrlein, ‘A comparative analysis of measures of social homogeneity’ (1987) 21 Quality and Quantity 219-31.
- Xeuming Lou, Siliang Tong, Zheng Fang and Zhe Qu ‘Fontiers Machines vs. Humans: The Impact of Artificial Intelligence Chatbot Disclosure on Customer Purchases’ (2019) 38(6) Marketing Science 937-947.

- Xizhu Xiao, Porismita Borah and Yan Su, 'The dangers of blind trust: Examining the interplay among social media news use, misinformation identification, and news trust on conspiracy beliefs' (2021) 30(8) Public Understanding of Science 977-992.
- Yanmengqian Zhou and Lijiang Shen, 'Confirmation Bias and the Persistence of Misinformation on Climate Change' (2021) Communications Research 1-24.
- Yazan Boshmaf, Ildar Muslukhov, Konstantin Beznosov and Matei Ripeanu, 'The socialbot network: when bots socialize for fame and money' (2011) Proceedings of the 27th Annual Computer Security Applications Conference 93-102.
- Zach Bastick, 'Would you notice if fake news changed your behavior? An experiment on the unconscious effects of disinformation' (2021) 116 Computers in Human Behavior <https://www.sciencedirect.com/science/article/pii/S0747563220303800> accessed 13 June 2023.
- Zeynep Tufekci, 'Engineering the public: big data, surveillance and computational politics' (2014) 19(7) First Monday <https://firstmonday.org/ojs/index.php/fm/article/view/4901/4097> accessed 13 June 2023.

Chapters in Edited Books

- Ethan Zuckerman, 'Intermediary Censorship' in Ronald Deibert, John Palfrey, Rafal Rohozinski and Jonathan Zittrain, *Access Controlled: The Shaping of Power, Rights and Rule in Cyberspace* (The MIT Press 2010).
- Jacques Herbots, 'Interpretation of contracts' in Jan Smits (ed), *Elgar Encyclopedia of Comparative* Law (Edward Elgar 2006).
- James Gordley, 'The Functional Method' in Pier Giuseppe Monateri (ed), *Methods of Comparative Law* (Edward Elgar 2012).
- Jeffrey J. Strange, 'How Fictional Tales Wag Real-World Beliefs' in Melanie C. Green, Jeffrey J. Strange and Timothy C. Brock, *Narrative Impact: Social and Cognitive Foundations* (2002 Taylor & Francis Group).

- Jerry Kang, ‘Bits of Bias’ in Justin D. Levinson and Robert J. Smith (eds.), *Communications Law* (2012 Cambridge University Press).
- Lee Ross and Andrew Ward, ‘Naïve Realism in Everyday Life: Implications for Social Conflict and Misunderstanding’ in Edward Reed, Elliot Turiel and Terrance Brown (eds), *Values and Knowledge* (Psychology Press 1996).
- Lee Ross, ‘Reactive Devaluation in Negotiation and Conflict Resolution’ in Kenneth Arrow, Robert Mnookin, Lee Ross, Amos Tversky, and Robert Wilson, *Barriers to Conflict Resolution* (W W Norton & Company 2007).
- Lee Ross, Mark Lepper and Andrew Ward, ‘History of Social Psychology: Insights, Challenges, and Contributions to Theory and Application’ in Susan Fiske, Daniel Gilbert, and Gardner Lindzey, *Handbook of Social Psychology* (Vol.1, 5th edn Wiley 2010).
- Mike Hajimichael, ‘Social Memes and Depictions of Refugees in the EU: Challenging Irrationality and Misinformation with Media Literacy Intervention’ in Alison MacKenzie, Jennifer Rose and Ibrar Bhatt, *The Epistemology of Deceit in a Postdigital Era: Dupery by Design* (2021 Springer).
- Nilanjana Dasgupta, ‘Implicit Attitudes and Beliefs Adapt to Situations: A Decade of Research on the Malleability of Implicit Prejudice, Stereotypes and the Self-Concept’ in Patricia Devine, Ashby Plant (eds., Vol 47), *Advances in Experimental Social Psychology* (2013 Academic Press).
- Vicki Jackson, ‘Comparative Constitutional Law: Methodologies’ in Michael Rosenfeld and András Sajó (eds), *The Oxford Handbook of Comparative Constitutional Law* (Oxford University Press 2012).

Books

- Alexander Meiklejohn, Free Speech and Its Relation to Self-Government (Harper & Brothers 1948).

- Alvin Gouldner, *The Dialectic of Ideology and Technology: The Origins, Grammar and Future of Ideology* (1976).
- Cass R. Sunstein, *Behavioral Law and Economics* (2012 Cambridge University Press).
- Cass R. Sunstein, *Behavioral Law and Economics* (2012 Cambridge University Press).
- Eric Barendt, *Freedom of Speech* (Oxford University Press 2007).
- Frederick Schauer, *Free Speech: A Philosophical Enquiry* (Cambridge University Press 1982).
- Geoffrey Samuel, *An Introduction to Comparative Law Theory and Method* (Hart Publishing 2014).
- Hannah Arendt, *Between Past and Future* (Viking Press 1968).
- Henry Jenkins, Sam Ford and Joshua Green, *Spreadable Media* (2013 New York University Press).
- Jeff Kosseff, *The United States of Anonymous: How the First Amendment Shaped Online Speech* (Cornell University Press 2022).
- John Milton, *Areopagitica with a Commentary by Sir Richard C. Jebb and With Supplementary Material* (1918 Cambridge University Press).
- John Stuart Mill, *On Liberty* (Yale University Press 2003).
- Jürgen Habermas (trs), *The Theory of Communicative Action* (1987).
- Konrad Zweigert and Hein Kötz, *An Introduction to Comparative Law* (Tony Weir tr, 3rd edn, Oxford University Press 1998).
- Kristin Lord, The Perils and Promise of Global Transparency: Why the Information Revolution May Not Lead to Security, Democracy or Peace (State University of New York Press 2006).
- Leontin-Jean Constantinesco, *Traité de Droit Comparé, Tome II: La Méthode Comparative* (Libraire Générale de Droit et de Jurisprudence 1974).
- Limor Shifman, *Memes in Digital Culture* (2014 MIT Press).
- Richard Dawkins, *The Selfish Gene: 40th Anniversary Edition* (2016 OUP).

- Richard J. Gerrig, *Experiencing narrative worlds: On the psychological activities of reading* (1993 Yale University Press).
- Richard Posner, *Economic Analysis of Law* (7th ed, Aspen Publishers 2007).
- Stanley Fish, *There's No Such Thing As Free Speech: And It's a Good Thing, Too* (Oxford University Press 1994).
- Valerie Steele, *The Corset – A Cultural History* (Yale University Press 2003).
- Zeynep Tufekci, 'Twitter and Tear Gas: The Power and Fragility of Networked Protest' (Yale University Press 2017).

Hansard

- HL Deb 18 May 2021.
- HC Deb 19 April 2022.
- HC Deb 11 May 2022.
- HL Deb 12 May 2022.

Government Reports

- Briefing Paper, Regulating Online Harms, Number 8743, 28 May 2021.
- Cabinet Office, Transparency in digital campaigning: Government response (2021) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993905/Digital_imprints_-_FINAL_Government_consultation_response_.pdf> accessed 16 November 2022.
- Cabinet Office, Transparency in digital campaigning: Technical consultation on digital imprints (2020) <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1014665/Digital_imprints_consultation.pdf> accessed 16 November 2022.
- Democracy and Digital Technologies Committee, *Digital Technology and the Resurrection of Trust* (HL 2019-21, 77).

- Department for Digital, Culture, Media & Sport, *Government Response to the Report of the Joint Committee on the Draft Online Safety Bill* (CP640, 2022).
- Department for Digital, Culture, Media and Sport, *Online Harms* (CP 57, 2019).
- Department for Digital, Culture, Media and Sport, *Online Harms White Paper: Full Government Response to the Consultation* (CP 354, December 2020).
- House of Commons committee – Digital, Culture, Media and Sport Committee, *Disinformation and 'fake news': Interim Report* (HC 2017-19, 363).
- Joint Committee on the Draft Online Safety Bill, *Draft Online Safety Bill* (2021-22, HL 129, HC 609).
- Joint Committee on the Draft Online Safety Bill, *Draft Online Safety Bill* (2021-22, HL 129, HC 609).
- Law Commission and Scottish Law Commission, 'Electoral Law: A joint final report' (Law Com No 389, Scot Law Com No 256, 2020).
- Law Commission, *Modernising Communications Offences* (Law Com No 399, 2021).
- PBC (Bill 209) 13 December 2022.

Other

- @artbot_ <https://twitter.com/artbot__> accessed 25 January 2022.
- @censusAmericans <https://twitter.com/censusamericans> accessed 25 January 2022.
- @emilydicknsnbot, <https://twitter.com/emilydicknsnbot> accessed 25 January 2022.
- @greatartbot <https://twitter.com/greatartbot> accessed 25 January 2022.
- @oliviataters <https://twitter.com/oliviataters> accessed 25 January 2022.
- @poem_exe <https://twitter.com/poem_exe> accessed 25 January 2022.
- '£350 million EU claim "a clear misuse of official statistics"' <https://fullfact.org/europe/350-million-week-boris-johnson-statistics-authority-misuse/> accessed 17 April 2023.

- ‘Austin man says sorry for posting misleading anti-Trump protester Tweet’ (13 November 2016, FOX 29 Philadelphia) <https://www.fox29.com/news/austin-man-says-sorry-for-posting-misleading-anti-trump-protester-tweet> accessed 20 April 2023.
- ‘Effective prevention and treatment for all respiratory viruses including Covid and Influence’ (Doctor Myhill, July 2022) <https://www.drmyhill.co.uk/wiki/Effective_prevention_and_treatment_for_all_respiratory_viruses_including_Covid_and_Influenza> accessed 23 April 2022.
- ‘FALSE: A video where dolphins appear to be swimming in a marina, supposedly due to the inactivity of the port caused by the coronavirus health crisis. This video has been circulated saying that it is the Promenade of Palma de Mallorca, the port of Denia (Alicante), the port of Moraira (Alicante) or the port of Premià de Mar (Barcelona).’ (Poynter., 18 April 2020) <https://www.poynter.org/?ifcn_misinformation=a-video-where-dolphins-appear-swimming-in-a-marina-supposedly-due-to-the-inactivity-of-the-port-caused-by-the-coronavirus-health-crisis-this-video-has-been-circulated-saying-that-it-is-the-promenade> accessed 13 June 2023.
- ‘meme, n.’ (*OED Online*, OUP April 2023) <https://www.oed.com/view/Entry/239909> accessed 8 April 2023.
- Alanna McKnight, ‘The Kurious Kase of Kim Kardashian’s Korset’ (2020) 3(1) Fashion Studies <https://www.fashionstudies.ca/the-kurious-kase-of-kim-kardashians-korset> accessed 13 June 2023.
- Alex Dalbey, ‘Trans people keep getting suspended from Twitter–and they want answers (updated)’ (daily dot, 21 May 2021) <https://www.dailydot.com/irl/trans-twitter-bans/> accessed 2 May 2023.
- Alex Hern, ‘Cambridge Analytica did work for Leave.EU, emails confirm’ <https://www.theguardian.com/uk-news/2019/jul/30/cambridge-analytica-did-work-for-leave-eu-emails-confirm> accessed 22 April 2022.

- Alex Hern, ‘Facebook agrees to pay fine over Cambridge Analytica scandal’ (30 October 2019, The Guardian) <https://www.theguardian.com/technology/2019/oct/30/facebook-agrees-to-pay-fine-over-cambridge-analytica-scandal> accessed 16 November 2022.
- Alistair Clark, ‘Elections Bill: a modest proposal to improve the Speaker’s Committee on the Electoral Commission’ <https://consoc.org.uk/elections-bill-a-modest-proposal-to-improve-the-speakers-committee-on-the-electoral-commission/> accessed 30 September 2022.
- Amnesty International, ‘A Human Rights Approach to Tackle Disinformation: Submission to the Office of the High Commissioner for Human Rights’ (14 April 2022) 12-3 <https://www.amnesty.org/en/wp-content/uploads/2022/04/IOR4054862022ENGLISH.pdf> accessed 17 April 2023.
- Anita Bhadani, ‘Democracy fears following “authoritarian” grab of Electoral Commission’ (The National, 28 April 2022) <https://www.thenational.scot/news/20101860.democracy-fears-following-authoritarian-grab-electoral-commission/> accessed 30 September 2022.
- Barbara Mikkelson & David Mikkelson, ‘Did Cher Have Ribs Removed To Make Her Waist Smaller?’ (Snopes, 22 June 2000) <https://www.snopes.com/fact-check/getting-waisted/> accessed 13 June 2023.
- BBC News, ‘Cambridge Analytica “not involved” in Brexit referendum, says watchdog’ <https://www.bbc.co.uk/news/uk-politics-54457407> accessed 22 April 2022.
- Berta García-Orosa, Pablo Gamallo, Patricia Martín-Rodilla and Rodrigo Martínez-Castaño, ‘Hybrid Intelligence Strategies for Identifying, Classifying and Analyzing Political Bots’ (2021) 10(10) Social Sciences <https://www.mdpi.com/2076-0760/10/10/357> accessed 13 June 2023.
- Bradley Smith, ‘Misnamed “Honest Ads Act” would restrict free speech’ (USA Today, 12 June 2019) <https://eu.usatoday.com/story/opinion/2019/06/12/election-interference-honest-ads-act-threatens-free-speech-editorials-debates/1438271001/> accessed 27 August 2022.
- Brian Contreras, ‘“I need my girlfriend off TikTok”: How hackers game abuse-reporting systems’ (Los Angeles Times, 3 December 2021)

<https://www.latimes.com/business/technology/story/2021-12-03/inside-tiktoks-mass-reporting-problem> accessed 2 May 2023.

- Caitlin Dewey, 'Facebook fake-news writer: "I think Donald Trump is in the White House because of me"' (2016 Washington Post) <https://www.washingtonpost.com/news/the-intersect/wp/2016/11/17/facebook-fake-news-writer-i-think-donald-trump-is-in-the-white-house-because-of-me/> accessed 12 April 2023.
- Council of Europe, Committee of Ministers, Recommendation CM/Rec(94)13 on measures to promote media transparency (22/11/1994), 'General provisions on media transparency.'
- Council of Europe, *Information Disorder: Towards an interdisciplinary framework for research and policy making* (September 2017), available at <https://rm.coe.int/information-disorder-report-version-august-2018/16808c9c77> accessed 26 September 2022.
- Craig Silverman & Lawrence Alexander, 'How Teens in The Balkans Are
- Craig Silverman, 'This Analysis Shows How Viral Fake Elections News Stories Outperformed Real News on Facebook' (Buzzfeed News, 16 November 2016) <https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook> accessed 22 April 2022.
- David Ingram, 'Facebook says 126 million Americans may have seen Russia-linked political posts' (Reuters, 30 October 2017) <https://www.reuters.com/article/us-usa-trump-russia-socialmedia/facebook-says-126-million-americans-may-have-seen-russia-linked-political-posts-idUSKBN1CZ2OI> accessed 22 April 2022.
- Digital Information World, 'Nearly 80% of Social Media Users have Adjusted their Privacy Settings in the Last Year' <https://www.digitalinformationworld.com/2019/10/research-shows-internet-users-taking-action-on-privacy.html#:~:text=Around%2079.2%20percent%20of%20the,profiles%20due%20to%20privacy%20concerns.> accessed 15 April 2023.
- Dominic Cummings, 'On the referendum #20: the campaign, physics and data science – Vote Leave's 'Voter Intention Collection System' (VICS) now available for all'

<https://dominiccummings.com/2016/10/29/on-the-referendum-20-the-campaign-physics-and-data-science-vote-leaves-voter-intention-collection-system-vics-now-available-for-all/> accessed 5 March 2022.

- Doug Zanger, 'Industry Opinion: Is the Honest Ads Act a viable solution for digital political advertising?' (The Drum, 24 October 2017) <https://www.thedrum.com/news/2017/10/24/industry-opinion-the-honest-ads-act-viable-solution-digital-political-advertising> accessed 27 August 2022.
  Duping Trump Supporters with Fake News' (Buzzfeed, 3 November 2016) <https://www.buzzfeednews.com/article/craigsilverman/how-macedonia-became-a-global-hub-for-pro-trump-misinfo> accessed 19 April 2023..
- Electoral Commission, 'Digital campaigning: Increasing transparency for voters' (June 2018) <https://www.electoralcommission.org.uk/sites/default/files/pdf_file/Digital-campaigning-improving-transparency-for-voters.pdf> accessed 18 October 2022.
- Emily Cleary, 'GP suspended after pushing vitamins and ivermectin to treat COVID' (Yahoo!News, 6 February 2023) <https://uk.news.yahoo.com/gp-suspended-after-pushing-vitamins-and-ivermectin-to-treat-covid-172806333.html> accessed 23 April 2023.
- Erin Griffith, 'Pro-Gun Russian Bots Flood Twitter After Parkland Shooting' (WIRED, 15 February 2018) <https://www.wired.com/story/pro-gun-russian-bots-flood-twitter-after-parkland-shooting/> accessed 16 April 2023.
- Facebook, 'Terms of Service' <https://m.facebook.com/terms/> accessed 1 May 2023.
- George Trefgarne, 'The Spirit of Orwell lives – that's the Ministry of Truth of it' *The Telegraph* (London, 16 July 2001) <https://www.telegraph.co.uk/finance/2726243/The-spirit-of-Orwell-lives-thats-the-Ministry-of-Truth-of-it.html> accessed 1 June 2023.
- Google search <https://www.google.com/search?q=site%3Atumblr.com+marilyn+manson+got+the+bottom+half+of+his+ribcage+removed> accessed 13 June 2023.

- Gordon Pennycook, Ziv Epstein, Mohsen Mosleh, Antonio A. Arechar, Dean Eckles & David G. Rand, 'Shifting attention to accuracy can reduce misinformation online' (2021) 592(7855) Nature, available at <https://www.nature.com/articles/s41586-021-03344-2> accessed 10 April 2023.
- ICO, 'Investigation into data analytics for political purposes' <https://ico.org.uk/action-weve-taken/investigation-into-data-analytics-for-political-purposes> accessed 16 November 2022.
- ICO, 'Investigation into the use of data analytics in political campaigns: A report to Parliament' (6 November 2018) <https://ico.org.uk/media/action-weve-taken/2260271/investigation-into-the-use-of-data-analytics-in-political-campaigns-final-20181105.pdf> accessed 16 November 2022.
- ICO, 'Legislation we cover' <https://ico.org.uk/about-the-ico/what-we-do/legislation-we-cover/> accessed 16 November 2022.
- ICO, 'Microtargeting' <https://ico.org.uk/your-data-matters/be-data-aware/social-media-privacy-settings/microtargeting/> accessed 18 October 2022.
- Internet users made up 7% of the world population in 2000 and 60% in 2020 – The World Bank, 'Individual using the Internet (% of population)' <https://data.worldbank.org/indicator/it.net.user.zs> accessed 23 April 2022.
- Jacob Rowbottom, 'Cakes, Gay Marriage and the Right against Compelled Speech' (UK Constitutional Law Association, 16 October 2018) <https://ukconstitutionallaw.org/2018/10/16/jacob-rowbottom-cakes-gay-marriage-and-the-right-against-compelled-speech/> accessed 18 October 2022.
- Jane Mayer, 'How Russia Helped Swing the Election for Trump' *The New Yorker* (New York City, 24 September 2018) <https://www.newyorker.com/magazine/2018/10/01/how-russia-helped-to-swing-the-election-for-trump> accessed 22 April 2022.
- Jim Waterson, 'UK fines Facebook £500,000 for failing to protect user data' (25 October 2018, The Guardian) <https://www.theguardian.com/technology/2018/oct/25/facebook-fined-uk-privacy-access-user-data-cambridge-analytica> accessed 16 November 2022.

- Josh Lowe, ‘Michael Gove: I’m “Glad” Economic Bodies Don’t Back Brexit’ <https://www.newsweek.com/michael-gove-sky-news-brexit-economics-imf-466365> accessed 22 April 2022.
- Josie O’Brien, ‘ABSOLUTE LIES: I was a toxic fitness influencer – here’s the lies I told and why you shouldn’t buy into ab workouts for a flat belly’ (The Sun, 4 January 2023) <https://www.thesun.co.uk/fabulous/20936909/toxic-fitness-influencer-abs-lies/> accessed 23 April 2023.
- Kevin Foster, ‘Press release: Government safeguards UK elections’ (Gov.uk, 5 May 2019) <https://www.gov.uk/government/news/government-safeguards-uk-elections> accessed 18 October 2022.
- Lachlan Markay and Andrew Desiderio, ‘How Gridlock, Social Media Giants and the Clintons Made the Internet Ripe for Russian Meddling’ (The Daily Beast, 20 October 2017) <https://www.thedailybeast.com/how-gridlock-social-media-titans-and-the-clintons-turned-the-internet-into-the-wild-west-of-american-politics> accessed 4 May 2023.
- Lee Goodman, ‘“Honest” political ads: Watch out, Drudge, you’re next’ (The Hill, 4 September 2019) <https://thehill.com/opinion/cybersecurity/459896-honest-political-ads-watch-out-drudge-youre-next/> accessed 27 August 2022.
- Margi Murphy, ‘U.S. Plan to Track Misinformation Sparks Its Own Misinformation’ (Bloomberg, 11 May 2022) accessed 1 June 2023.
- Marília Gehrke, ‘Transparency as a key element of data journalism: perceptions of Brazilian professionals’ (Computation + Journalism Symposium, 2020) available at <https://cpb-us-w2.wpmucdn.com/sites.northeastern.edu/dist/d/53/files/2020/02/CJ_2020_paper_8.pdf> accessed 2 November 2022.
- Mark Steyn, ‘The Show Ofcom Won’t Let You See’ (Steyn Online, 16 March 2023) <https://www.steynonline.com/13331/the-show-ofcom-wont-let-you-see> accessed 1 June 2023.

- Matthew Kelly, ‘Before Trump, Cambridge Analytica was on team Cruz’ <https://www.opensecrets.org/news/2018/03/before-trump-cambridge-analytica-was-on-team-cruz/> accessed 22 April 2022.
- Max de Haldevang, ‘Russian trolls and bots are flooding Twitter with Ford-Kavanaugh disinformation’ (Quartz, 2 October 2018) <https://qz.com/1409102/russian-trolls-and-bots-are-flooding-twitter-with-ford-kavanaugh-disinformation> accessed 16 April 2023.
- Meta, ‘Misinformation’ <https://transparency.fb.com/policies/community-standards/misinformation> accessed 1 May 2023.
- Michael Gove saying “People in this country have had enough of experts.” See David Matthews, ‘Brexit would be a victory for those who distrust academics’ <https://www.timeshighereducation.com/blog/brexit-would-be-victory-those-who-distrust-academics> accessed 22 April 2022.
- Michela Palese and Josiah Mortimer, ‘Reigning in the Political “Wild West”: Campaign Rules for the 21st Century (Electoral Reform Society, February 2019) <https://www.electoral-reform.org.uk/latest-news-and-research/publications/reining-in-the-political-wild-west-campaign-rules-for-the-21st-century/> accessed 18 October 2022.
- Nick Cohen, ‘The Tories call it electoral reform. Looks more like a bid to rig the system.’ (Guardian, 18 December 2021) <https://www.theguardian.com/commentisfree/2021/dec/18/the-tories-call-it-electoral-reform-looks-more-like-a-bid-to-rig-the-system> accessed 30 September 2022.
- Ofcom, ‘News Consumption in the UK: 2022’ (21 July 2022) 2, available at <https://www.ofcom.org.uk/__data/assets/pdf_file/0024/241827/News-Consumption-in-the-UK-Overview-of-findings-2022.pdf> accessed 20 April 2022.
- Office of the United Nations High Commissioner for Human Rights, ‘New and emerging technologies need urgent oversight and robust transparency: UN experts’ (2 June 2023) <https://www.ohchr.org/en/press-releases/2023/06/new-and-emerging-technologies-need-urgent-oversight-and-robust-transparency> accessed 3 June 2023.

- Oxford English Dictionary, ‘algorithm, n.’ <https://www-oed-com.ezphost.dur.ac.uk/view/Entry/4959?redirectedFrom=algorithm#eid> accessed 21 April 2023.
- Peter Hess, ‘The User Experience Researcher Who Was Fooled by a Twitter Bot’ <https://www.vice.com/en/article/kb7bmn/olivia-taters-twitter-bots> accessed 25 January 2022.
- Rachit Shukla, Adwitiya Sinha and Ankit Chaudhary, ‘TweezBot: An AI-Driven Online Media Bot Identification Algorithm for Twitter Social Networks’ (2022) 11(5) Electronics <https://www.mdpi.com/2079-9292/11/5/743> accessed 13 June 2023.
- RaeAnn Christensen, ‘Austin man says sorry for posting misleading anti-Trump protester Tweet’ (FOX 4 News, 14 November 2016) <https://www.fox4news.com/news/austin-man-says-sorry-for-posting-misleading-anti-trump-protester-tweet> accessed 3 June 2023.
- Rasmus Kleis Nielsen and Lucas Graves, ‘“News you don’t believe”: Audience perspectives on fake news’ (Reuters Institute for the Study of Journalism 2017) <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-10/Nielsen%26Graves_factsheet_1710v3_FINAL_download.pdf> accessed 14 April 2023.
- Recommendation CM/Rec(2012)4 of the Committee of Ministers to member states on the protection of human rights with regard to social networking services (4 April 2012).
- Reuters, ‘What are the links between Cambridge Analytica and a Brexit campaign group?’ <https://www.reuters.com/article/us-facebook-cambridge-analytica-leave-eu-idUSKBN1GX2IO> accessed 22 April 2022.
- Sawdah Bhaimiya, ‘Several left-wing activists had their Twitter accounts suspended after a false-report campaign by far-right users’ (INSIDER, 1 December 2022) <https://www.businessinsider.com/left-wing-activists-banned-from-twitter-after-false-report-2022-11?r=US&IR=T> accessed 2 May 2023.
- Scott Shane and Vindu Goel, ‘Fake Russian Facebook Accounts Bought $100,000 in Political Ads’ The New York Times (New York City, 6 September 2017)

<https://www.nytimes.com/2017/09/06/technology/facebook-russian-political-ads.html> accessed 22 April 2022.

- See Alex Green, 'Mark Steyn show on GB News breached Ofcom code with Covid claims' (The Independent, 6 March 2023) <https://www.independent.co.uk/news/uk/ofcom-gb-news-naomi-wolf-b2294926.html> accessed 1 June 2023.
- Statista, 'Discord brand awareness, usage, popularity, loyalty, and buzz among messenger users in the UK in 2022.' <https://www.statista.com/forecasts/1328633/discord-messengers-brand-profile-in-the-uk> accessed 13 June 2023.
- Statista, 'Steps taken by global internet users to protect online activities and personal information as of December 2022' <https://www.statista.com/statistics/617422/online-privacy-measures-worldwide/> accessed 15 April 2023.
- The Electoral Commission, 'A strategy and policy statement for the Electoral Commission' (5 July 2021) <https://www.electoralcommission.org.uk/who-we-are-and-what-we-do/our-views-and-research/elections-act/a-strategy-and-policy-statement-electoral-commission> accessed 30 September 2022.
- The Electoral Commission, 'Know who is paying for online political ads' <https://www.electoralcommission.org.uk/i-am-a/voter/online-campaigning/know-who-paying-online-political-ads> accessed 23 April 2023.
- The Electoral Commission, 'Political Finance Regulation and Digital Campaigning: A Public Perspective : GfK UK report for qualitative research findings' 24 April 2018.
- The Electoral Commission, 'The Electoral Commission's ability to bring prosecutions' (5 July 2021) <https://www.electoralcommission.org.uk/who-we-are-and-what-we-do/our-views-and-research/elections-act/electoral-commissions-ability-bring-prosecutions> accessed 30 September 2022.
- The Electoral Commission, 'Transparency in digital campaigning: response to Cabinet Office technical consultation on digital imprints' available at <https://www.electoralcommission.org.uk/who-we-are-and-what-we-do/changing-electoral-

law/transparent-digital-campaigning/transparency-digital-campaigning-response-cabinet-office-technical-consultation-digital-imprints> accessed 16 November 2022.

- tumblr user 'powerburial' <https://powerburial.tumblr.com/post/88245733235/popunklouis-remember-that-rumor-we-all-believed> accessed 13 June 2023.
- Twitter, 'How we address misinformation on Twitter' <https://help.twitter.com/en/resources/addressing-misleading-info> accessed 1 May 2023.
- Twitter, Twitter User Agreement <https://cdn.cms-twdigitalassets.com/content/dam/legal-twitter/site-assets/privacy-policy-new/Privacy-Policy-Terms-of-Service_EN.pdf> accessed 1 May 2023.
- Wikitionary, 'do not feed the troll' <https://en.wiktionary.org/wiki/don%27t_feed_the_troll> accessed 21 April 2023.
- World Health Organisation, 'Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation' <https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation> accessed 14 April 2023.
- Ziv Epstein, Adam J. Berinsky, Rocky Cole, Andrew Gully, Gordon Pennycook, and David G. Rand, 'Developing an accuracy-prompt toolkit to reduce COVID-19 misinformation online' (2021) 2(3) Harvard Kennedy School Misinformation Review <https://misinforeview.hks.harvard.edu/article/developing-an-accuracy-prompt-toolkit-to-reduce-covid-19-misinformation-online/> accessed 10 April 2023.