

Tackling Online False Information in the United Kingdom

Coe, Peter

DOI:

[10.1080/17577632.2024.2316360](https://doi.org/10.1080/17577632.2024.2316360)

License:

Creative Commons: Attribution-NonCommercial-NoDerivs (CC BY-NC-ND)

Document Version

Publisher's PDF, also known as Version of record

Citation for published version (Harvard):

Coe, P 2024, 'Tackling Online False Information in the United Kingdom: The Online Safety Act 2023 and its Disconnection from Free Speech Law and Theory', *Journal of Media Law*.
<https://doi.org/10.1080/17577632.2024.2316360>

[Link to publication on Research at Birmingham portal](#)

General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact UBIRA@lists.bham.ac.uk providing details and we will remove access to the work immediately and investigate.

Tackling online false information in the United Kingdom: The Online Safety Act 2023 and its disconnection from free speech law and theory*

Peter Coe

To cite this article: Peter Coe (20 Feb 2024): Tackling online false information in the United Kingdom: The Online Safety Act 2023 and its disconnection from free speech law and theory*, *Journal of Media Law*. DOI: 10.1080/17577632.2024.2316360

To link to this article: <https://doi.org/10.1080/17577632.2024.2316360>




© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 20 Feb 2024.



Submit your article to this journal 

[View related articles](#) View Crossmark data 

Tackling online false information in the United Kingdom: The Online Safety Act 2023 and its disconnection from free speech law and theory*

Peter Coe

Birmingham Law School, University of Birmingham, Birmingham, UK

ABSTRACT

It is commonly recognised that the publication of false information can be harmful to the public sphere. The Online Safety Act 2023 places statutory responsibilities on regulated services to prevent the publication of certain false information. This article interrogates the regime's compatibility with established free speech law and theory. I argue that there is a disconnect between the legislation and the legal and theoretical principles underpinning free speech, which could have insidious and long-lasting implications for the right and the public sphere.

ARTICLE HISTORY Received 29 September 2023; Accepted 21 December 2023

KEYWORDS Online Safety Act 2023; free speech; free speech theory; online speech; false information

Introduction

In the UK, there has been consistent recognition from a variety of actors, including the UK government, that the dissemination of false information can be harmful to individuals and the public sphere.¹ It has also been acknowledged that this problem is being exacerbated by the role played in our lives by the likes of Google, Facebook, Instagram, and X,² and because

CONTACT Peter Coe  p.j.coe@bham.ac.uk

*This article was presented as a paper at 'The Regulation of Disinformation: A Critical Appraisal Workshop' at Goethe University Frankfurt am Main, Germany, 7th–8th September 2023, and at the Institute of Advanced Study, Durham University, 22nd January 2024. I am indebted to Rebecca Moosavian (University of Leeds), Dr Bosko Tripkovic (University of Birmingham), Professor András Koltay (Pázmány Péter Catholic University and University of Public Service, Budapest), Dr Eliza Bechtold (University of Aberdeen) and the *Journal of Media Law*'s anonymous reviewer for their invaluable feedback on previous drafts of this article.

¹Department for Digital, Culture, Media and Sport, *Online Harms White Paper: Full Government Response to the Consultation* (CP 354, 2020), [34], 84–85.

²For example, see: J Bayer, I Katsirea et al, European Parliament, 'The Fight against Disinformation and the Right to Freedom of Expression', July 2021; P Coe, 'The Draft Online Safety Bill and the Regulation of Hate Speech: Have We Opened Pandora's Box?' (2022) 14 *Journal of Media Law* 50, 51.

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

the systems that were in place for dealing with this type of content (and other illegal and/or harmful content), prior to the introduction of the Online Safety Act 2023 (OSA), were designed for the offline world, and were (and in some cases, still are) outdated and no longer fit for purpose.³

The UK's online harms regime has intensified this debate. The regime began life in April 2019 as the Online Harms White Paper,⁴ morphing into multiple iterations of the Online Safety Bill (OSB), published in its original form in May 2021, and finally crystallising as the OSA, which was enacted on the 26th of October 2023. On the one hand, it is acknowledged that legislation placing statutory responsibilities on internet services to prevent the publication of false information (and other illegal and harmful content) may benefit society and public discourse.⁵ This is because, in theory at least, by helping to decrease the volume of false information we are exposed to, such laws should reduce the opportunities for the public sphere to become distorted. As citizens we should be able to assess, with greater confidence, the veracity of information available to us, and in turn, use this information, and the trust we have in it, to make positive contributions to public discourse.⁶

But, on the other hand, the OSA has been (and before it, the OSB was) met with significant resistance from a variety of actors because of the potential threats to free speech that it presents.⁷ Indeed, since the publication of the White Paper, and the initial draft of the OSB, the regime has been shrouded in controversy. The OSB was subject to numerous amendments, and at one stage, it looked as though it would be scrapped altogether. Yet despite this, at the time of writing, the OSA has recently been enacted, albeit the overall shape of the regime remains unclear, because much of the legal detail will be contained in secondary legislation. Therefore, debates on the efficacy of the OSA will continue, and only time will tell what its ultimate impact on free speech will be.⁸

Notwithstanding this uncertainty, the purpose of this article is to interrogate the regime's compatibility with free speech law and theory. In doing so, it begins with an explanation of what is meant by false information, and how the phenomenon has been exacerbated by the internet. This is followed by analysis of the pre-OSA system for dealing with this content, and an

³Dame Melanie Dawes, Ofcom, 'In News We Trust: Keeping Faith in the Future of Media' (Oxford Media Convention, 19 July 2021) (Keynote speech).

⁴HM Government, *Online Harms White Paper* (CP 57, 2019).

⁵Coe (n 2) 51.

⁶P Coe, *Media Freedom in the Age of Citizen Journalism* (Edward Elgar 2021), 74–85.

⁷For example, see House of Lords Communications and Digital Committee, *Free for All? Freedom of Expression in the Digital Age* (HL Paper 54, 22 July 2021); M Earp, 'UK Online Safety Bill Raises Censorship Concerns and Questions on Future of Encryption', Committee to Protect Journalists, 25 May 2021; L Kirkconnell-Kawana, 'Online Safety Bill: Five Thoughts on its Impact on Journalism' Media@LSE, 3 June 2021; C Elsom, 'Safety without Censorship. A Better Way to Tackle Online Harms' Centre for Policy Studies, September 2020.

⁸Coe (n 2) 51. According to Ofcom's current roadmap to regulation, the regulator will adopt a phased approach to the OSA's implementation: <<https://www.ofcom.org.uk/online-safety/information-for-industry/roadmap-to-regulation/0623-update>> accessed 14 December 2023.

explanation of why it did not work, as aspects of it have a bearing upon the OSA regime. Next, the contours of the free speech framework are sketched, including relevant jurisprudence of the European Court of Human Rights (ECtHR), and the theories underpinning it that are particularly relevant to online false information. In this section I explain why these theories are flawed in this context, and therefore how these flaws could justify the creation of laws to tackle online false information. Yet, as I go on to suggest in my analysis of the OSA, which follows, this creates a paradoxical disconnect between theory and law, in that although the flaws in the theories may justify the creation of such laws – which manifests as the OSA – its creation arguably conflicts with the ECtHR’s jurisprudence, and the spirit of its theoretical foundations, and could inadvertently interfere with free speech. Finally, the article concludes with some potential solutions for meeting this challenge that do not erode one of the core fundamental human rights.

Contextualising online ‘false information’

In the UK the terms disinformation, misinformation, malinformation, fake news, false information, and false news are referred to ubiquitously, and often, mistakenly, interchangeably, by a variety of state and private actors,⁹ relating to a wide range of topics and debates that are, predominantly, associated with modern technology and online communication. It is, therefore, important to make two points explicit at the outset of this article relating to, firstly, what is meant by false information, and secondly, how the internet, social media and the press contribute to the false information phenomenon.

What is meant by false information?

Fake news, false information, and false news, are generic terms that are applied to either or both of disinformation and misinformation, and, less commonly, ‘malinformation’.¹⁰ In this article, when referring to the distinct species of disinformation, misinformation, and malinformation I will use the umbrella term ‘false information’.

Misinformation *tends* to refer to information that is fully or partially incorrect but is spread without intending to deceive the recipients of the

⁹The vagueness and incorrect application of the terminology is well documented in literature: T McGonagle, ‘“Fake News”: False Fears or Real Concerns?’ (2017) 35 Netherlands Quarterly of Human Rights 203, 203–09; T Venturini, ‘Confession of a FakeNews Scholar’, (2018) 68th Annual Conference – International Communication Association, Prague; E Shattock, ‘Fake News in Strasbourg: Electoral Disinformation and Freedom of Expression in the European Court of Human Rights’ (2022) 13(1) European Journal of Law and Technology, 4–5; E C Tandoc Jr. et al, ‘Defining “Fake News”’ (2018) 6(2) Digital Journalism, 137–53.

¹⁰Council of Europe, ‘Dealing with Propaganda, Misinformation and Fake News’: <<https://www.coe.int/en/web/campaign-free-to-speak-safe-to-learn/dealing-with-propaganda-misinformation-and-fake-news>> accessed 11 December 2023.

information.¹¹ Disinformation is more sinister and insidious, as it refers to untrue information that is purposefully crafted and strategically placed for the purpose of deceiving the recipient(s) into believing a lie or taking action that, for instance, serves political or commercial interests.¹² Unhelpfully, at times, misinformation is conflated with another, distinct, species of false information, as demonstrated by evidence submitted to the Law Commission in respect of its *Modernising Communications Offences report*.¹³ Consultees such as Full Fact and Demos suggested that misinformation can also capture *true* information, or information with some elements of truth, that is deliberately used in a misleading way,¹⁴ with the intent to cause harm rather than serve the public interest. Describing misinformation in this way is problematic, as the distinction between misinformation (as it tends to be understood) and disinformation lies in the purpose of the publisher of the content. As I discuss later in this article, this is important in the context of imposing liability under existing law and under the OSA regime. Misinformation as articulated by Full Fact and Demos is actually a distinct form of false information that should be referred to as ‘mal-information’.¹⁵ As I suggest later, in my analysis of the OSA regime, the confusion created by conflating two distinct species of false information – misinformation with malinformation – could contribute to a liability gap in the new regime.

The role played by the internet, social media and the press in the false information phenomenon

Although false information tends to be associated with the internet and social media, its manipulative properties have been exploited by those in power, such as monarchs, the church, and state and private actors, for centuries.¹⁶ For instance, and as I will return to in respect of the OSA, it is, and always has been, synonymous with our tabloid press. Throughout its history press barons have used propaganda to advance their own agendas,¹⁷ and up to the present day, the lucrative trade in celebrity gossip in our tabloid press provides an example of the use of what is often untrue, or only partially true, information for financial gain.¹⁸

¹¹ *Online Harms White Paper* (n 4), 23.

¹² P N Howard, *Lie Machines* (Yale University Press 2020), 15.

¹³ Law Commission, *Modernising Communications Offences: A Final Report* (HC 547, Law Com 399, 2021).

¹⁴ *Ibid.* According to Full Fact misinformation is ‘often deliberately designed to be not false but to create a false impression’ and that it ‘is often simple to manipulate a false claim into a true claim that is in effect misleading’ [3.44], 86; See also, Demos’s submission at [3.27], 81.

¹⁵ Council of Europe (n 10).

¹⁶ P Bernal, ‘Fakebook: Why Facebook Makes the Fake News Problem Inevitable’ (2018) 69(4) *Northern Ireland Legal Quarterly*, 513, 516–19; *The Internet, Warts and All* (Cambridge University Press 2018), chp. 9; I Cram, ‘Keeping the Demos Out of Liberal Democracy? Participatory Politics, ‘Fake News’ and the Online Speaker’ (2019) 11(2) *Journal of Media Law* 113, 129.

¹⁷ T Driberg, *Beaverbrook: A Study in Power and Frustration* (Weidenfeld & Nicolson 1956), 213.

¹⁸ Coe (n 6), 191–192; J Oster, *Media Freedom as a Fundamental Right* (Cambridge University Press, 2015), 38–39; for detailed commentary on press malfeasance generally, see: P Wragg, *A Free and Regulated*

The difference today is that the internet and social media act as a false information Petri dish, by providing the ideal technological architecture and environment for false information to grow and spread quickly, to potentially millions of people, both online and offline.¹⁹ Social media's ability to do this is amplified by the symbiotic relationship that now exists between online content and factions of our media, as it is often used as a source of news. The fact that 'trusted' mainstream media publish what may be false information, serves to justify and support that false information, thereby creating a self-fulfilling and insidious cycle.²⁰ This situation has not been helped in recent years by the state of the press industry, which has led to an almost perma-state of 'hyperactivity'.²¹ This 'faster and shallower corporate journalism', which necessitates the need for newspapers to provide news 24 hours-a-day across multiple platforms, combined with fewer journalists, and an increasing reliance on clickbait and sensationalist headlines to generate clicks and advertising revenue,²² has encouraged churnalism,²³ which leads, in some cases, to 'fast and loose' journalism that sees professional values and appropriate source and fact checking being cast aside.²⁴ Ultimately, this results in more mistakes,²⁵ including the inadvertent dissemination of false information.²⁶

The way in which news is presented on social media also inadvertently helps to spread false information. The *Cairncross Review*, which is one of the most important and extensive studies conducted on the sustainability of UK journalism, found that 'fake news ... is particularly hard to spot on social media, where news content is often presented alongside content that has no relationship to news at all'²⁷ and that consequently 'online consumption makes it harder for public-interest news to reach audiences, but easier for fake news to do so',²⁸ which ultimately makes it harder for audiences to discern falsity. From a UK perspective, this is contributing to a decline

Press: Defending Coercive Independent Press Regulation (Hart Publishing 2020). From the ECtHR, see: *Mosley v United Kingdom* App. no. 48009/08 (ECtHR 10 May 2011), [114]; *Von Hannover v Germany* (No 1) App. no. 59320/00 (ECtHR 24 June 2004), [65]; *Hachette Filipacchi Associes v France* App. no. 12268/03 (ECtHR 23 July 2009), [40]; *Eerikainen and others v Finland* App. no. 3514/02 (ECtHR 10 February 2009), [62]; *Standard Verlags GmbH v Austria* (No 2) App. no. 21277/05 (ECtHR 4 June 2009), [52]; *MGN Ltd v United Kingdom* App. no. 39401/04 (ECtHR 18 January 2011), [143].

¹⁹Coe (n 6), 81–85; A Bruns et al, 'When a Virus Goes Viral: Pros and Cons to the Coronavirus Spread on Social Media' *Inform*, 22nd March 2020.

²⁰*ibid.* (Coe, Bruns); P Coe, 'The Good, The Bad and The Ugly of Social Media during the Coronavirus Pandemic' (2020) 25(3) *Communications Law* 119, 119–22.

²¹Coe (n 6) 68.

²²*ibid.* 70–72.

²³N Fenton, 'Regulation Is Freedom: Phone Hacking, Press Regulation and the Leveson Inquiry – the Story so far' (2018) 23(3) *Communications Law* 118, 119.

²⁴*ibid.*

²⁵R L Weaver, *From Gutenberg to the Internet: Free Speech, Advancing Technology, and the Implications for Democracy* (2nd edn, Carolina Academic Press 2019), 202.

²⁶Coe (n 6) 68; Cram (n 16) 129.

²⁷The Cairncross Review, *A Sustainable Future for Journalism* (12th February 2019) 33.

²⁸*ibid.*

in the trust we have in our press industry.²⁹ The recently published King's College London's World Values Survey found that of the UK citizens surveyed only 13 per cent said they trusted the press – the second lowest ranking of the 24 countries surveyed – and for Generation Z,³⁰ this falls to 5 per cent.³¹ As a consequence, people are turning away from newspapers to other sources of news, and to predominantly online platforms and services.³² This metamorphosis in our news consumption habits, that is animated by the latest Ofcom News Consumption report, tells us that in the UK, for adults aged over 16, only 26 per cent use the print version of newspapers (increasing to 39 per cent if the online platform is included), whereas 68 per cent use the internet for their news.³³ For 16–24-year-olds, of the top 4 news sources, 3 are social media platforms (Instagram, X and TikTok), with *BBC One* being the only non-social media source.³⁴

From a public sphere perspective, the problem with this is that false information online is not just an issue that is fused to the press. There have been many other high-profile examples of the internet and social media contributing to the problem without the press being involved. Cambridge Analytica, for instance, harvested over 50 million user profiles without Facebook's permission and manufactured sex scandals and disinformation to influence voters in elections globally, including the UK,³⁵ and *The Guardian* has revealed the extent of the disinformation-for-profit market, in which private contractors, employed by companies and politicians, have used social media to manipulate elections worldwide – a practice that it is being predicted to continue during the forthcoming UK and US elections.³⁶

Thus, the Oxford University Reuters Institute Digital News Report 2023 found that 56 per cent of the 96,000 people surveyed worldwide worry about identifying what news is real and false online, and for those who say they mainly use social media as a source of news the figure rises to 64 per

²⁹Ipsos Veracity Index 2022 <<https://www.ipsos.com/en-uk/ipsos-veracity-index-2022>>.

³⁰People born from 1997 onwards: M Dimock, 'Defining Generations: Where Millennials End and Generation Z Begins' (Pew Research Centre, 17th January 2019).

³¹Egypt has the lowest ranking at 8 per cent: King's College London, *World Values Survey March 2023*. See also: Statista, Share of adults who trust news media most of the time in selected countries worldwide as of February 2023: <https://www.statista.com/statistics/308468/importance-brand-journalist-creating-trust-news/> accessed 14th December 2023.

³²Coe (n 6) 69.

³³Ofcom, *News Consumption in the UK: 2023* (July 2023), 3–4.

³⁴*ibid* 14.

³⁵Howard (n 12), 12.

³⁶S Kirchgaessner et al, 'Revealed: The Hacking and Disinformation Team Meddling in Elections', *The Guardian*, 15th February 2023. See also: D Milmo and A Hern, 'Elections in UK and US at Risk from AI-driven Disinformation, say experts', *The Guardian*, 20th May 2023. See also: House of Commons Digital, Culture Media and Sport Committee, *Disinformation and 'Fake News': Final Report* (HC 1791, 18 February 2019), 68–77; For a global perspective on this issue, see generally: S Bradshaw and P N Howard, *The Global Disinformation Order 2019 Global Inventory of Organised Social Media Manipulation* (Oxford Internet Institute and University of Oxford 2019).

cent,³⁷ which is concerning when you bear in mind the news consumption trend discussed above.³⁸ Accordingly, Philip Howard sums up the impact that this type of content has on the media, public sphere and democracy generally:

While the internet has certainly opened new avenues for civic participation in political processes – inspiring hopes of democratic reinvigoration ... divisive social media [false news] campaigns have heightened ethnic tension, revived nationalistic tensions, intensified political conflict, and even resulted in political crisis – while simultaneously weakening public trust in journalism, democratic institutions and electoral outcomes.³⁹

False information is, therefore, a multi-industry and supranational problem that has, in recent years, contributed to the acute pressure that governments around the world are being put under to sanitise the online environment – with the UK government being no exception. This is largely based on the increasingly popular notion that despite the benefits to free speech and the public sphere wrought by online services, their role in proliferating and intensifying harmful content warrants a rethink of their contribution to society and democracy, as well as their motives and responsibilities. This brings me to the next section of this article; that is how false information in the UK was dealt with under the pre-OSA regime. Here, I will briefly set out the system and its problems. This is important because aspects of it will continue to run concurrently with the OSA regime and/or will feed directly into the new regime.

The pre-Online Safety Act 2023 system

In the UK the pre-OSA system for tackling online false information can be separated into three distinct components, that I split into hard law, soft law, and quasi or non-legal responses to the problem.

Component one relies on hard law to impose legal liability on the individual or the publisher responsible for the false information *that is in some way illegal*, because for example, it is defamatory, breaches privacy law, copyright law, or data protection law, or as is most common, a criminal offence. The problems with imposing criminal liability on individuals responsible for publishing false information animate how difficult the situation is, both under the pre-OSA system, and, as discussed later in this article, under the OSA regime.

Communications offences within UK criminal law are contained within a suite of offences that were not designed with online communication in mind. Section 127(2) of the Communications Act 2003, for instance, deals with

³⁷Reuters Institute, University of Oxford, *Digital News Report 2023*, 17.

³⁸Of those surveyed in the UK, the most common false information categories were politics, climate change, the war in Ukraine and, even now, Covid. Ibid.

³⁹Howard (n 12) 18.

false messages.⁴⁰ The offence applies to situations where a person sends a message, or causes a message to be sent, over a public communications network, for the purpose of causing annoyance, inconvenience or needless anxiety to another⁴¹; and they know the message to be false.⁴² Thus, it must be established that the defendant's purpose was to cause annoyance, anxiety or needless anxiety to another, and that the defendant knew, rather than believed, the communication to be false.⁴³ Setting aside the number of publishers that could theoretically be prosecuted for publishing such content, which in itself is resource-intensive, the transience of online publishers, the fact they can be located and/or operate in different jurisdictions, and the frequency with which they publish anonymously or pseudonymously, means that locating and identifying them is challenging.⁴⁴ This is compounded for prosecutors by having to prove beyond reasonable doubt the defendant's knowledge of falsity, which in itself can be complex and is an evidence-intensive task.⁴⁵ Ultimately, this has led to under-criminalisation,⁴⁶ with the Law Commission noting that the section 127(2) offence is 'infrequently prosecuted'.⁴⁷

Component two is relying on soft law, in the hope that platforms will sign up to and adhere to voluntary, non-binding co-regulatory or self-regulatory codes of conduct, such as the European Commission's Code of Practice on Disinformation.⁴⁸ Unfortunately, it seems that these codes have not worked.⁴⁹ In its own assessment, the European Commission found a number of shortcomings in the Code, such as the inconsistent and incomplete application of the relevant commitments from services and member states, and, more broadly, the insufficient scope of the Code (in that it

⁴⁰As does section 1(1)(a)(iii) of the Malicious Communications Act 1988. Analysing both the 2003 and 1988 Acts, and their respective false information offences, is beyond the scope of this article. I focus on section 127(2) because this is the offence which is most commonly engaged in relation to social media: D McGoldrick, 'The Limits of Freedom of Expression on Facebook and Social Networking Sites: A UK Perspective,' (2013) 13 Human Rights Law Review 125, 132 citing D Ormerod, 'Telecommunications: Sending Grossly Offensive Message By Means of a Public Electronic Communications Network' (2007), Jan, Criminal Law Review, 98–100.

⁴¹Communications Act 2003 127(2)(a)–(b).

⁴²*ibid* 127(2)(a).

⁴³*ibid*. See also: Crown Prosecution Service, 'Social Media and other Electronic Communications', Legal Guidance, 9th January 2023.

⁴⁴Coe (n 2) 57–58.

⁴⁵Law Commission (n 13), [3.25], 81.

⁴⁶*ibid*. [1.5], 2.

⁴⁷*ibid*. [3.13], 78.

⁴⁸This was updated in 2022 in line with the European Union's Digital Services Act. This Act came into force on the 25th of August 2023 for very large online platforms, such as X and Facebook. It becomes fully applicable to other entities on the 17th of February 2024. The UK will not be subject to it due to our exit from the European Union.

⁴⁹The same can be said for the EU Code of Conduct on Hate Speech. See: Coe (n 2) 56–60; T Quintel and C Ullrich, 'Self-regulation of Fundamental Rights?' The EU Code of Conduct on Hate Speech, related initiatives and beyond in B Petkova and T Ojanen (eds), *Fundamental Rights Protection Online. The Future Regulation of Intermediaries* (Edward Elgar 2021) 197–229.

does not address several issues of the online ecosystem).⁵⁰ Moreover, services have used the Code for self-serving interests, namely, to enhance, or preserve, reputations, and to avoid more direct and onerous regulatory oversight.⁵¹ This has meant that services sign up to Codes to simply pay lip service to them.

Component three is relying on quasi-legal and non-legal responses. This includes services own internal policies on false information and other harmful content. However, like self-regulatory codes of conduct, it seems that services will often devise such policies for self-serving purposes, and then rarely apply them to tackle the problem in practice.⁵² This component also includes support provided by, for example, NGOs, charities, civil society, the education sector, and press regulators. For instance, in the past, grants have been given to the UK fact-checking charity Full Fact, and to the independent body First Draft, which offer guidance to journalists on verifying content on social media.⁵³ In a similar vein, the Guidance to the new Impress Journalism Standards Code provides extensive advice to its regulated members on how to ensure accuracy, and limit the potential for spreading false information, when using online sources.⁵⁴ However, although there is a lot of commendable and valuable work going on within this component, it tends to be reliant on a mixture of funding, goodwill, and the buy-in of those affected, therefore its effectiveness can be inconsistent, and difficult to assess with accuracy.⁵⁵

This ‘mish-mash’ of legal, quasi-legal and non-legal responses to online false information represents a piecemeal analogue-based regime that is undoubtedly flawed, and is not able to adequately cope with the scale, scope and nuances of illegal and harmful false information published online; a situation compounded by the fact that platforms are simply not doing enough to tackle the problems internally, and in many cases, they seem to be ignoring it completely.⁵⁶

The OSA, which I return to after the following section, presents ways of dealing with this problem (and online harms generally). There I give an overview of the regime, and analyse how it may tackle online false information, and in doing so, I explain how it creates a disconnect with established free

⁵⁰European Commission, *Assessment of the Code of Practice on Disinformation – Achievements and Areas for Further Improvement* (Commission Staff Working Document) swd(2020) 180 final 7–19; P Cavaliere, ‘The Truth in Fake News: How Disinformation Laws Are Reframing the Concepts of Truth and Accuracy on Digital Platforms’ (2022) 3(4) *European Convention on Human Rights Law Review* 481, section 3.

⁵¹Shattock (n 9) 3.

⁵²Dawes (n 3).

⁵³Coe (n 20) 121.

⁵⁴Impress Standards Code and Guidance, clause 1: <<https://www.impressorg.com/wp-content/uploads/2023/02/Impress-Standards-Code.pdf#page=15>>.

⁵⁵Bernal, *Warts and All* (n 16) 247–48.

⁵⁶Dawes (n 3).

speech theories and doctrine. Thus, in the next section, I briefly sketch the free speech law and theory that bear on this debate.

The free speech framework: jurisprudence and theory

Section 2 of the Human Rights Act 1998 (HRA 1998) requires domestic judges to take ECtHR jurisprudence into account in domestic proceedings, an obligation that has previously been interpreted strictly by the House of Lords.⁵⁷ Consequently, domestic case law should ‘mirror’ the jurisprudence of the ECtHR.⁵⁸ According to Lord Bingham in *R (on the application of Ullah) v Special Adjudicator*,⁵⁹ failure to follow ‘clear and constant’ Strasbourg jurisprudence would be unlawful under section 6(1) HRA 1998,⁶⁰ unless there are ‘special circumstances’⁶¹ that justify departure from that approach.⁶² Similarly, section 3 of the 1998 Act imposes an obligation on the judiciary to interpret legislation in conformity with Article 10.

Article 10(1) of the European Convention on Human Rights (ECHR) protects freedom of expression by providing that: ‘Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers.’ Article 10(2) qualifies this right, in that a state can restrict the Article 10(1) right in the interests of, inter alia, ‘the prevention of disorder or crime, the protection of health or morals, the protection of the reputation or rights of others’. The ECtHR’s jurisprudence provides that the scope of protection under Article 10 is broad, in that it covers information or ideas that are unpalatable to the state, or offending or shocking to some people,⁶³ and the exceptions to which it is subject, are interpreted narrowly by the Court. In line with this broad conception of the right, the protection under Article 10 extends to the sharing of information that is strongly suspected to be untruthful,⁶⁴

⁵⁷In *R (on the application of Ullah) v Special Adjudicator* [2004] 2 AC 323, 350 Lord Bingham stated that the ‘duty of the national courts is to keep pace with the Strasbourg jurisprudence as it evolves over time: no more, but certainly no less’.

⁵⁸*R (on the application of Quark Fishing Ltd) v. Secretary of State for Foreign and Commonwealth Affairs* 1 AC 529, [2006], 34 per Lord Nicholls. See also: P Wragg, ‘A Freedom to Criticise? Evaluating the Public Interest in Celebrity Gossip after *Mosley and Terry*’ (2010) 2(2) *Journal of Media Law* 295, 314.

⁵⁹[2004] 2 AC 323.

⁶⁰Section 6(1) states: ‘[I]t is unlawful for a public authority to act in a way which is incompatible with a Convention right’; pursuant to section 6(3), the definition of ‘public authority’ includes the courts.

⁶¹It is unclear what amounts to ‘special circumstances’: see Wragg (n 58) 314.

⁶²*ibid.*

⁶³*Handyside v United Kingdom* App no 5493/72 (ECHR, 7 December 1976) [49]. See also: *Sunday Times v United Kingdom* (No. 1) App no 6538/74 (ECHR, 26 April 1979) [65]; *Lingens v Austria* App no 9815/82 (ECHR, 8 July 1986) [41]; *Axel Springer AG v Germany* (No. 1) App no 39954/08 (ECHR, 7 February 2012) [78]; *Thorgeir Thorgeirson v Iceland* App no 13778/88 (ECHR, 25 June 1992) [63].

⁶⁴D J Harris et al, *Law of the European Convention on Human Rights* (2nd edn, Oxford University Press, 2009) 444–45; In *Salov v Ukraine* App. No. 65518/01, 6 September 2005, the Court found, at [13]: ‘Article 10 of the Convention as such does not prohibit discussion or dissemination of information received even if it is strongly suspected that this information might not be truthful. To suggest otherwise would deprive persons of the right to express their views and opinions about statements made in

meaning that laws that generally prohibit the dissemination of false information merely on the ground of its falsity, without regard for additional factors such as the harm caused to personal rights, are likely to contravene Article 10(1).⁶⁵ The Court's Article 10 jurisprudence is based on theories of speech that are distinct, yet are intertwined and complementary,⁶⁶ in that they encompass different aspects of the right.⁶⁷ However, despite their relevance to ECtHR jurisprudence, when it comes to online speech, particularly in the context of online false information, for the reasons I explain below, these theories – particularly those concerned chiefly with truth-discovery, which I turn to first – are flawed. This creates a paradoxical disconnect between theory and law, in that the flaws in the theories underpinning the Court's free speech case law arguably justify the creation of laws to tackle false information. However, as I suggest in the remainder of this article, the creation of such laws conflict with the ECtHR's jurisprudence, and the spirit of its theoretical foundations, and could inadvertently interfere with free speech.

Argument from truth and the marketplace of ideas

Of relevance to the publication of online false information is John Stuart Mill's argument from truth,⁶⁸ and the marketplace of ideas, which broadly accord with the Court's jurisprudence on laws that restrict the dissemination of false information merely on the basis of falsity.⁶⁹

The argument from truth is concerned with 'epistemic advance'.⁷⁰ Indeed, Mill regards truth, at times, as merely a by-product of open discussion.⁷¹ Of paramount importance to Mill is not the discovery of truth, but the process

the mass media and would thus place an unreasonable restriction on the freedom of expression set forth in Article 10 of the Convention.'

⁶⁵ J Hoboken et al., 'The legal framework on the dissemination of disinformation through internet services and the regulation of political advertising', IVIR, December 2019, 39; I Katsirea, 'Fake News: Reconsidering the Value of Untruthful Expression in the Face of Regulatory Uncertainty' (2018) 10(2) *Journal of Media Law* 159, 171–76; Bayer (n 2) 24, 26.

⁶⁶ *R v Secretary of State for the Home Department, ex parte Simms* [2000] 2 AC 115, per Lord Steyn, 126. See also: *R (on the application of Lord Carlisle of Berriew QC and others) v Secretary of State for the Home Department* [2014] UKSC 60 per Lord Kerr, [164].

⁶⁷ J Oster, *European and International Media Law* (Cambridge University Press, 2017), 41; V Blasi, 'The Checking Value in First Amendment Theory' (1977) *American Bar Foundation Research Journal* 521, 554; E Barendt, *Freedom of Speech* (2nd edn, Oxford University Press, 2005) 6–7; P Wragg, 'Mill's Dead Dogma: The Value of Truth to Free Speech Jurisprudence' (2013), Apr, *Public Law* 363.

⁶⁸ J S Mill, *On Liberty, Essays on Politics and Society* in J.M. Robson (ed), *Collected Works of John Stuart Mill*, vol. XVIII (University of Toronto Press, 1977) ch. 2, 228–59.

⁶⁹ *Salov* (n 64) [13]; As stated below, the argument from democratic self-governance is chief among the theories supporting the ECtHR's free speech jurisprudence. However, it has been argued that libertarianism remains the de facto communication theory for online speech in Western democracies, and the two theories that predominantly underpin libertarianism are the argument from truth and the marketplace of ideas. See P Coe, '(Re)embracing Social Responsibility Theory as a Basis for Media Speech: Shifting the Normative Paradigm for a Modern Media' (2018) 69(4) *Northern Ireland Legal Quarterly*, 403, 406.

⁷⁰ F Schauer, *Free Speech: A Philosophical Enquiry* (Cambridge University Press 1982) 25.

⁷¹ J Gray, *Mill on Liberty: A Defence* (2nd edn, Routledge 1996) 110.

of discussion and debate.⁷² Mill argues that the foundations and reasoning upon which opinions are based must be continually tested and, as result, the acceptance of alternative views by others, and ultimately the reliable discovery of truth, must derive from effective persuasion, rather than coercion.⁷³ Additionally, Mill says that why we should not use truth to determine what is acceptable and unacceptable speech, and therefore, by extension, why we should not regulate based on truth, has four facets.⁷⁴ Firstly, the state would expose its own fallibility if it suppresses opinion on account of that opinion's perceived falsity as, in fact, it may be true.⁷⁵ Secondly, even if the suppressed opinion is objectively false, it has some value, as it may (and in Mill's opinion very commonly does) contain an element of truth.⁷⁶ Thirdly, since the dominant opinion on any given subject is rarely, or never, the whole truth, what remains will only appear as a result of the collision of adverse opinions.⁷⁷ Finally, notwithstanding the third facet, even if the received opinion is not only true, but the entire truth, unless it is rigorously discussed and debated, it will not carry the same weight, as the rationale behind it may not be fully and accurately comprehended.⁷⁸ Consequently, unless opinions can be frequently and freely challenged, by forcing those holding them to defend their views, the very meaning and essence of that true belief may, itself, be weakened, become ineffective, or even lost⁷⁹: In Mill's words, the true belief: 'will be held as a dead dogma, not a living truth'.⁸⁰

In his philosophical enquiry into free speech Frederick Schauer suggested that the desirability of truth within society is almost universally accepted⁸¹ – a statement that is difficult to argue against. However, in reality, the assumption that is often drawn from Mill's argument, that free speech leads to truth, can be attacked on the following, related, grounds when it is applied to online speech. Firstly, there is not, necessarily, a causal link between free speech and the discovery of truth.⁸² This is particularly relevant to online speech, where anybody disseminate information, meaning that the internet is saturated with content that is inaccurate, misleading, or untrue. In other words, online speech is not always conducive to revealing 'truth'. Secondly, online

⁷²Schauer (n 70) 20.

⁷³Mill (n 68) 217–23.

⁷⁴For analysis of this aspect of Mill's argument, see: C MacLeod, 'Mill on the Liberty of Thought and Discussion' in A Stone and F Schauer (eds), *The Oxford Handbook of Freedom of Speech* (Oxford University Press 2021) 3–19, 8–10.

⁷⁵Mill (n 68) 258; See generally: Barendt (n 67) 8.

⁷⁶*ibid* (Mill) 229.

⁷⁷*ibid* 252, 258.

⁷⁸*ibid* 258.

⁷⁹*ibid* 258; See also: Wragg (n 18) 139–40; Wragg (n 67) 365.

⁸⁰*ibid* (Mill) 243.

⁸¹Schauer (n 70) 26.

⁸²*ibid* 15.

false information continues to contribute to entrenched political polarisation.⁸³ This often occurs in echo chambers, or filter bubbles, in which political opinions and prejudices are fostered through insular engagement with like-minded people, which is the opposite of what Mill argued for.⁸⁴

Although the marketplace of ideas is a distinct theory, it is regarded as deriving from Mill's argument from truth.⁸⁵ It comes *Abrams v United States*,⁸⁶ in which Justice Holmes said: 'the best test of truth is the power of the thought to get itself accepted in the competition of the market'.⁸⁷ Thus, the theory is based on the premise that 'truth', or the 'best' ideas, will win out, as they will naturally emerge from the competition of ideas in the marketplace.⁸⁸ According to Eric Barendt the theory 'rests on shaky grounds'⁸⁹: an infirmness that is exposed by the naivety of the theory in the context of the online speech marketplace for the following reasons: Firstly, if the assertion that one statement is stronger than another cannot be intellectually supported and defended, the notion of truth loses its integrity,⁹⁰ as history demonstrates: falsehood frequently triumphs over truth, to the detriment of society.⁹¹ Secondly, in line with Jürgen Habermas's concept of discourse and the public sphere, which aims at reaching a rationally motivated consensus and is based on the assumption of the prevalence of reason,⁹² the theory assumes that recipients of the content consider what they read or view within the context of the speech marketplace rationally; deciding whether to accept or reject it, based on whether it will improve their lifestyle, and society generally.⁹³ However, as stated above in respect of the argument from truth, the internet proliferates a huge amount of

⁸³Coe (n 6) 79–85, 151–53.

⁸⁴*ibid.* See also: K Klionick, 'The New Governors: The People, Rules and Processes Governing Online Speech' (2018) 131 *Harvard Law Review* 1599, 1665; A Koltay, *New Media and Freedom of Expression Rethinking the Constitutional Foundations of the Public Sphere* (Hart Publishing 2019), 199; N Stroud, 'Media Use and Political Predispositions: Revisiting the Concept of Selective Exposure' (2008) 30 *Political Behaviour* 341–65.

⁸⁵In *Simms* (n 66) 126, Lord Steyn treated Mill's argument from truth and Justice Holmes's marketplace of ideas as interchangeable. This view is supported by commentators such as Schauer (n 70) 15–16. *cf.* Coe (n 6) 130–31; Wragg (n 67) 368–69, V Blasi, 'Reading Holmes through the Lens of Schauer' (1997) 72(5) *Notre Dame Law Review* 1343, 1355, Barendt (n 67) 11–13, who treat the theory as a distinct interpretation, or form, of the argument from truth.

⁸⁶250 US 616 (1919).

⁸⁷250 US 616 (1919), 630–31.

⁸⁸*ibid.*; See also *Gitlow v New York* 268 US 652 (1925), 673 per Justice Holmes.

⁸⁹Barendt (n 67) 12; E Barendt, 'The First Amendment and the Media' in I Loveland (ed), *Importing the First Amendment: Freedom of Speech and Expression in Britain, Europe and the USA* (Hart Publishing 1998) 43–46; J Weinberg, 'Broadcasting and Speech' (1993) 81 *California Law Review* 1103, 1162.

⁹⁰Barendt (n 67) 12.

⁹¹R Abel, *Speech and Respect* (Stevens & Sons Limited 1994) 48; D Milo, *Defamation and Freedom of Speech* (Oxford University Press 2008) 57.

⁹²J Habermas, *The Structural Transformation of the Public Sphere* (Polity Press 1962); *The Theory of Communicative Action*, vol. 1: *Reason and the Rationalization of Society* (Beacon Press 1984), 25, 39, 99; *The Theory of Communicative Action*, vol. 2: *Lifeworld and System: A Critique of Functionalist Reason* (Beacon Press 1987), 120, 319; *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy* (William Rehg trans., Polity Press 1996).

⁹³Weinberg (n 89); J Skorupski, *John Stuart Mill* (Routledge 1991) 371–72.

information that is poorly researched or simply untrue, yet has the potential to, and very often does emerge as the dominant ‘view’ regardless of the detrimental impact this may have on individuals or society.⁹⁴ This issue is amplified by the ubiquity of anonymous and pseudonymous online speech, making it hard, if not impossible, for audiences to accurately and rationally assess the veracity of the speaker. Thus, in reality, in a marketplace that contains true and untrue or misleading information in at least equal proportions, some of which may be published anonymously or under a pseudonym, it may be impossible for recipients of the communication to make a rational assessment of what they have read or viewed.⁹⁵ Finally, for the reasons discussed above in the respect of the argument from truth, the marketplace of ideas’ basis of rationality is undermined by echo chambers and filter bubbles. Thus, within the context of online speech at least, as Jonathan Weinberg declares: ‘[t]o the extent that our most basic views and values are relatively immune to rational argument, the marketplace metaphor seems pointless’.⁹⁶

Argument from democratic self-governance

Despite the obvious relevance of the argument from truth and marketplace of ideas to false information, chief among the theories present in the Court’s jurisprudence is the argument from democratic self-governance,⁹⁷ which is based on the premise that the purpose of free speech is to protect the right of citizens to understand political matters in order to facilitate and enable societal engagement with the political and democratic process.⁹⁸ The argument was subsequently developed by Alexander Meiklejohn to encompass ‘public discourse’, which in essence dictates that free speech protects public discourse on all matters of public concern.⁹⁹ It is this expanded version of the argument that we see in the Court’s case law,¹⁰⁰ and in domestic case law.¹⁰¹

⁹⁴Coe (n 6) 151.

⁹⁵*ibid.*

⁹⁶Weinberg (n 89) 1162, 1159–160.

⁹⁷For example, see *Lingens v Austria* (1986) A 103, [42]; *Bladet Tromsø and Stensaa v Norway* (2000) 29 EHRR 125, [59]; *Bergens Tidende v Norway* (2001) 31 EHRR 16, [48]; *Thorgeir Thorgeirson v Iceland* App no 13778/88 (ECHR, 25 June 1992), [64]. Helen Fenwick and Gavin Phillipson observe that only the argument from democratic self-governance has been prominently employed by the ECtHR: H Fenwick and G Phillipson, *Media Freedom under the Human Rights Act* (Oxford University Press 2006) 39, 707–10.

⁹⁸R Bork ‘Neutral Principles and Some First Amendment Problems’ (1971) 47 *Indiana Law Journal* 1, 27–28; J Oster, ‘Theory and Doctrine of “Media Freedom” as a Legal Concept’ (2013) 5(1) *Journal of Media Law* 57, 69.

⁹⁹A Meiklejohn, *Political Freedom: The Constitutional Powers of the People* (Oxford University Press 1960) 42; A Meiklejohn, ‘The First Amendment is an Absolute’ [1961] *Supreme Court Review* 245, 255–257.

¹⁰⁰For example, see *Lingens v Austria* (1986) A 103, [42]; *Bladet Tromsø and Stensaa v Norway* (2000) 29 EHRR 125, [59]; *Bergens Tidende v Norway* (2001) 31 EHRR 16, [48]; *Thorgeir Thorgeirson v Iceland* App no 13778/88 (ECHR, 25 June 1992), [64].

¹⁰¹*Simms* (n 66), per Lord Steyn at 126; *Reynolds v Times Newspapers Limited* [2001] 2 AC 127 (HL) per Lord Cooke at 220; *Jameel v Wall Street Journal Europe Sprl* [2007] 1 AC 359 (HL) per Baroness Hale at [158].

The free speech framework: conclusion

As stated above, the flaws in these theories create a paradoxical disconnect between theory and law. The argument from truth, marketplace of ideas and, to a greater extent, the expanded version of the argument from democratic self-governance manifest in the social purpose of free speech:¹⁰² to enable citizens to engage in public discussion, and in doing so, positively contribute to the governance of their community.¹⁰³ On the one hand, clearly, the ubiquity of online false information, and the damage it causes undermines the assumption that often flows from the argument from truth and marketplace of ideas that open discussion and debate is the best way of furthering knowledge, and in doing so discovering truth and suppressing falsehoods. And, more broadly, false information has the potential to disrupt the social purpose of free speech, as it can distort our perception and understanding of matters of public concern, and erode the quality of debate (which Mill's argument from truth is so concerned about), thereby preventing us from engaging fully in the public sphere and with the democratic process.¹⁰⁴ However, as I argue in the following sections, on the other hand, the OSA could significantly interfere with the jurisprudence of the Court, and the essence of these theories. This could undermine our right to free speech and, domestically, give rise to a potential conflict with the HRA 1998 provisions, and the mirror principle, set out at the beginning of this section.

The Online Safety Act 2023 and online false information

Overview of the regime¹⁰⁵

The online safety regime, which began life with the Online Harms White Paper in 2019, found that the existing 'patchwork of regulation and voluntary initiatives' were not effective at keeping us safe online, and that online harms, in their many various forms, could only be tackled with the imposition of a single regulatory framework.¹⁰⁶ The White Paper developed into the OSB, and eventually, the OSA. At its core, the regime is risk-based, with regulated services required to conduct risk assessments of services regarding criminal content, as well as content harmful to children, and implement effective and proportionate risk-mitigation plans where harm arising from such content and/or the operation of their service is identified. Consequently, it was said the regime would 'end the era of self-

¹⁰²Oster (n 67) 40.

¹⁰³Habermas (n 92).

¹⁰⁴Coe (n 6) 81–85.

¹⁰⁵See OSA, s1.

¹⁰⁶Online Harms White Paper (n 4), 6 [7], 30.

regulation’,¹⁰⁷ by creating a regulatory system that imposes responsibility on the platforms themselves through statutory ‘hard and manifold’ safety duties of care¹⁰⁸ to protect users from certain illegal content, and in the case of children from some harmful and age-inappropriate content.¹⁰⁹ I describe these duties as ‘hard-edged’ as they *must* be met. In contrast, the free speech duties are ‘softer-edged’, as the regime’s language only requires services to ‘take account of’ or ‘have regard to’ them.¹¹⁰ As I discuss in the following section, this is significant because of what meeting the safety duties may mean for free speech when the ‘softer-edge’ of the free speech duties are taken into account.¹¹¹

Under the OSA, the Online Safety regulator is Ofcom, which has the power to fine companies up to £18 million, or 10 per cent of annual global turnover, whichever is higher, if they fail in their duty of care.¹¹² Furthermore: (i) Ofcom has the power to block non-compliant services from being accessed in the UK¹¹³; (ii) sections 144–148 provide for ‘business disruption measures’ that allow Ofcom to apply for a variety of ‘restriction orders’ if the regulated service has failed to meet certain conditions relevant to the restriction sought, and; (iii) section 110 creates criminal offences, pursuant to section 109, for named senior managers of in-scope services. Notwithstanding these powers, it is important to acknowledge here that Ofcom is *not* a censor, and therefore its role as a regulator is *not* to determine the acceptability or otherwise of individual pieces of content, nor is its role to remove content. Rather, its role is to ensure that regulated services have appropriate systems and processes in place to protect their users.

The OSA regulates providers of ‘internet services’.¹¹⁴ It separates these as user-to-user services, which incorporate typical social media platforms such as Facebook and Instagram etc.,¹¹⁵ search services, such as Google,¹¹⁶ and, under Part 5 of the Act, internet services displaying ‘regulated provider pornographic content’. It distinguishes regulated services further through a system of categorisation. Ofcom is under a duty to establish a register which will categorise services as either Category 1 user-to-user services only, Category 2A search services and user-to-user services which include a search engine, and Category 2B user-to-user services.¹¹⁷ Under the OSB,

¹⁰⁷Lord Bishop of Oxford, HL Deb 18th May 2021, vol. 812, col. 517.

¹⁰⁸Coe (n 2) 66.

¹⁰⁹In the OSA, the safety duties relating to adults are set out at sections 10 (user-to-user services) and 27 (search services) and for children at sections 12 and 29 (user-to-user and search services respectively).

¹¹⁰OSA, ss17,19,22.

¹¹¹Coe (n 2) 66.

¹¹²OSA, Sched 13, para 4.

¹¹³*ibid*, s144.

¹¹⁴*ibid*, s226.

¹¹⁵*ibid*, s3(1).

¹¹⁶*ibid*, s229.

¹¹⁷*ibid*, s95.

which Category a service fell into was to be determined by user numbers *and* functionality, as well as other factors the Secretary of State deemed to be relevant.¹¹⁸ However, because of an amendment to Schedule 11 and section 97 (4) of the OSB, which was included in the OSA,¹¹⁹ Ofcom is now required to consider functionality *independently* of user numbers when determining categorisation of a service. This means that we can say with some certainty that the likes of Facebook and X will fall into Category 1, but whether less popular platforms such as Reddit and Tumblr end up in Category 1 or 2B remain to be seen as the user number cut off between the two categories has not, yet, been clarified. In theory, the Schedule 11 amendment means that small, yet high-harm platforms, could be captured by the regime as ‘emerging Category 1 services’.¹²⁰ But, as this will be determined by Ofcom’s research into user numbers *and* functionality,¹²¹ which should happen within six months of enactment, but could be prolonged to up to eighteen months post-enactment,¹²² there is currently a lack of clarity as to which services could fall within section 97(4), which is compounded by another issue with how the regime will categorise regulated services. Schedule 11 requires the Secretary of State to make regulations specifying the Category 1, 2A and 2B threshold conditions, which will be set out in a post-enactment statutory instrument.¹²³ As I discuss in the next section, this is concerning because secondary legislation made outside of Parliament is not subject to the same rigorous, transparent, and publicly accessible parliamentary scrutiny as primary legislation, and is therefore arguably more susceptible to industry pressure and lobbying.

The OSA and false information: a difficult history

The amount of online false information generated about Covid was one of the key drivers behind the UK government initially including false information ‘that could cause significant harm to an individual’ within the scope of the regime as it was originally conceived, falling within its ‘legal but harmful’ provisions.¹²⁴ Furthermore, as explained above, a key aspect of the regime as it then stood was to end ineffective self-regulation.¹²⁵

However, rather controversially, the OSA goes back on these earlier statements of intent, for two reasons. Firstly, the legal but harmful provisions

¹¹⁸*ibid*, Sched 11, para 1.

¹¹⁹This amendment was tabled by Baroness Morgan of Cotes: <<https://bills.parliament.uk/bills/3137/stages/17765/amendments/96158>> accessed 13 December 2023.

¹²⁰OSA, s97.

¹²¹*ibid*, Sched 11, para 2.

¹²²*ibid*, Sched 11, para 2(10).

¹²³*ibid*, Sched 11, para 1 and s224(3).

¹²⁴DCMS (n 1), [34], 84–85; N Dorries, Statement UIN HCWS19: <<https://questions-statements.parliament.uk/written-statements/detail/2022-07-07/hcws194>> accessed 13 December 2023.

¹²⁵HL Deb (n 107).

relating to adults have been removed. This was because of concerns that in allowing what is legal but harmful to be determined by the Secretary of State, the category could be expanded in the future by virtue of secondary legislation which, as stated above in relation to the threshold conditions for categorising regulated services, is subject to less scrutiny than primary legislation, and that it could be used as a tool for repressive censorship, either by the state, or by regulated services, and that ultimately we would end up in a situation where content that is being removed online would still be legal offline.¹²⁶ However, as I will come back to in a moment, the OSA introduces a ‘triple shield’ – an aspect of the regime that I argue provides a back-door for the regulation of legal but harmful content. Secondly, as I discuss below, rather than ending an ‘era of self-regulation’, the OSA may result in self-regulation being put on a statutory footing.

Free speech concerns: a paradoxical disconnect between the Online Safety Act 2023 and free speech law and theory

The OSA tackles false information in different ways, including the creation of a new false communications offence and the ‘triple shield’.¹²⁷ It also provides controversial exemptions for ‘recognised news publishers’ in relation to the publication of false information. In this section, I discuss each of these aspects of the regime in turn as they give rise to considerable free speech concerns.

False communications offence

The false communications offence is based on recommendations made by the Law Commission in its *Modernising Communications Offences report*.¹²⁸ Pursuant to section 179 of the OSA, the offence is committed if (a) the person sends a message, (defined in section 182(2)-(3), as sending, transmitting or publishing a communication by electronic means, or causing such to happen), (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. Likely audience is defined as individuals for whom it is

¹²⁶For example, see: Carla Lockhart, MP for Upper Bann, HC Deb. 19th April 2022, vol. 712, col. 117.

¹²⁷Additionally, s150 OSA requires Ofcom to establish an Advisory Committee on disinformation and misinformation which, under s152(3)(a)-(c), must include persons representing the interests of users and regulated services, and persons with expertise in the prevention and handling of disinformation and misinformation online. Section 152(5) requires the Committee to public a report within the period of 18 months after being established, and after that must publish periodic reports. Thus, at the time of writing, the make-up, role, and influence of the Committee remain to be seen.

¹²⁸Law Commission (n 13), [2.38]-[2.39], 24.

reasonably foreseeable that they would encounter the message, or a subsequent message that forwards or shares the content of the original message.¹²⁹

The fault element of this offence has two aspects: (i) knowledge of falsity, in that the defendant knew, rather than believed, the message to be false – this is the same as the *mens rea* required for section 127(2) Communications Act 2003 offence¹³⁰ and (ii) the defendant intended the message to cause non-trivial psychological or physical harm. These do not operate in isolation, but rather must be taken together.¹³¹ This creates three issues.

Firstly, knowledge of falsity instantly limits the scope of the offence to disinformation only, which means that malinformation (that is, true information deliberately used in a misleading way) – it is not covered by the offence.¹³² This liability ‘gap’ in dealing with malinformation was raised by consultees to the Law Commission’s proposals for reform,¹³³ and was acknowledged by the Law Commission itself.¹³⁴ The Law Commission’s solution was to recommend the creation of a further ‘harm-based’ communications offence,¹³⁵ which was complete upon a defendant sending a message that was likely to cause harm to a likely audience, and in doing so, they intended to cause harm to the likely audience.¹³⁶ However, in the OSA, the harm-based offence has been removed as the government felt that its lower threshold posed a risk to free speech. On the one hand, it is arguable that the removal of the harm-based offence significantly limits the scope of the criminal regime, as it does not recognise the granularity of online communications.¹³⁷ But, on the other hand, as it stood, because of the inherent difficulties with defining and determining what is ‘harmful’ content and content that is ‘likely to cause harm’ (which was reflected in how earlier versions of the OSB described such content)¹³⁸ the offence would have likely fallen foul of the Article 10 principles I have set out above, in particular the ECtHR’s jurisprudence protecting speech that is unpalatable or even untruthful.¹³⁹ In turn, this could erode the social purpose of free speech, in that citizens would be denied exposure and access to content that may enable them to better engage in public discussion.¹⁴⁰

¹²⁹OSA, s179(2).

¹³⁰Law Commission (n 13), [3.19], 79.

¹³¹*ibid*, [3.55], 90.

¹³²*ibid*, [3.44], 86.

¹³³*ibid*, [3.43], [3.44], 86.

¹³⁴*ibid*, [3.47], 87.

¹³⁵*ibid*.

¹³⁶*ibid*, [2.38]–[2.39], 24.

¹³⁷A similar argument was made by Demos: *ibid*. [3.27], 81.

¹³⁸Coe (n 2) 20–21.

¹³⁹See (n 63–65).

¹⁴⁰See (n 102–104).

Secondly, and notwithstanding the point above, falsity itself is often difficult to ascertain.¹⁴¹ As Mill acknowledged in his argument from truth, with some content aspects of it may be true, partially true, or false. This will create significant challenges for the Crown Prosecution Service with proving knowledge of falsity. This complexity is compounded by the second aspect of the *mens rea* – establishing the intention of the defendant to cause non-trivial psychological or physical harm. We do not know, at the moment, what ‘non-trivial psychological or physical harm’ means. Indeed, Law Commission consultees made the point that it is hard to define,¹⁴² which may lead to a lack of clarity in the law. A further issue relates to a lack of clarity of the threshold of seriousness. The Law Commission was clear that to avoid over-criminalisation, by setting a low level of culpability, knowledge of falsity must be coupled with the intention to cause harm – which it says sets a higher threshold than ‘causing annoyance, inconvenience or needless anxiety’ under section 127(2) Communications Act 2003.¹⁴³ According to the Commission, setting the threshold at this level was critical to prevent a disproportionate interference with free speech¹⁴⁴ – which clearly accords with the free speech principles I have previously discussed. However, this creates two conflicting concerns. The first is that because the offence is linked to the intention to cause harm, but it does not refer to actual harm, there is a risk that the offence will be interpreted and applied over-broadly. To the contrary, the second is that the offence’s two-pronged *mens rea*, combined with the general difficulty in ascertaining falsity and non-trivial harm, and the current lack of clarity over the threshold, will make the offence difficult to prove, particularly in respect of borderline cases,¹⁴⁵ thereby potentially limiting the scope of the offence even further.

Finally, the same broad arguments apply to the operation of this offence as to the existing Communications Act 2003 offence that were advanced above – that is that the nature of online communication makes offences hard to prosecute. If anything, for the reasons advanced here, this new offence may make life for prosecutors even harder.

The ‘triple shield’

In essence, the ‘triple shield’ (i) requires all regulated services to swiftly remove all illegal content when notified of its existence, and to actively

¹⁴¹This point was made by English PEN: Law Commission (n 13) [3.25], 81.

¹⁴²*ibid*, [3.45], 86–87.

¹⁴³*ibid*, [3.54], 90; Harmful Online Communications: The Criminal Offences (2020) Law Commission Consultation Paper No 248, [6.45].

¹⁴⁴*ibid* (*Modernising Communications Offences*).

¹⁴⁵Law Commission (n 13) [3.53], 89–90.

monitor posts to prevent users from exposure to the worst material¹⁴⁶; (ii) requires Category 1 services to apply their terms of service consistently and fairly, and to remove content that is banned by their own terms of service, and; (iii) allows adults to tailor the type of content they see via toggles (in other words, the ability to switch between two different options), giving them the ability to potentially avoid harmful content should they not wish to see it, and giving them greater control over who they engage with.¹⁴⁷ It is the first two aspects of the shield which give rise to the greatest free speech concerns.

As stated, the OSA's safety duties require regulated services to swiftly remove 'illegal content'.¹⁴⁸ This is defined in section 59(2) as content amounting to a 'relevant offence',¹⁴⁹ which includes the false communications offence created by the OSA.¹⁵⁰ Section 192 says that services are required to find illegality if they have 'reasonable grounds to infer' that the elements of the offence – so, for example, the 'false communications' offence – are made out, including the *actus reus* and *mens rea* elements¹⁵¹ – and they do not have reasonable grounds to infer that a defence to the offence may be successfully relied upon.¹⁵² Thus, the critical issue is the intent of the user and whether the user has an available defence (such as a reasonable excuse). The problem is that unless the service has information on which it can infer that a defence may successfully be relied on, the possibility of a defence cannot be considered – the platform has to disregard it.¹⁵³ In other words, this is the first example of the OSA undermining the government's statement of intent to do-away with self-regulation, as this provision requires platforms to anticipate illegality – and therefore remove content – on the basis of information reasonably available to them, meaning that what could be valuable extrinsic contextual information is omitted from their assessment.

Ultimately, because of the sanctions regulated services are faced with, they are likely to err on the side of caution, and programme their algorithms that will manage this risk accordingly. This issue was raised by Graham Smith and Edina Harbinja in a policy report on the OSB, where they pointed out that this approach is likely to lead to the filtering and removal of legal online content at a scale that is incomparable to offline removal. And that

¹⁴⁶This first shield also requires regulated services to put measures in place to prevent their services being used for illegal activity, for instance.

¹⁴⁷Children will automatically have these settings by default – albeit how this will work in practice, and what, if any, duties this would impose on in-scope services is unclear.

¹⁴⁸OSA, s10(3)(b).

¹⁴⁹*ibid*, s59(4), (5): Content that is linked to priority or non-designated offences.

¹⁵⁰*ibid*, s59(4)(b), (5)(c).

¹⁵¹*ibid*, s192(5), (6)(a).

¹⁵²*ibid*, s192(6)(b).

¹⁵³*ibid*, s192(2), 192(6)(b).

the section 192 illegality duty is a form of prior restraint, as the regime requires content filtering and removal decisions to be made before any fully informed, fully argued decision on the merits takes place.¹⁵⁴ As a consequence, and in direct conflict with the ECtHR's jurisprudence against prohibiting the dissemination of false information merely on the ground of its falsity, regulated services may be forced into a position where they will have to do exactly that to protect their interests.

The second aspect of the triple shield relates to the duties imposed on Category 1 regulated services to take-down user-generated content that breach the regulated service's terms of service.¹⁵⁵ Additionally, Category 1 services must use '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 terms of service,¹⁵⁶ with 'relevant content' being content that the services terms of service state action will be taken against.¹⁵⁷ Clearly, this aspect of the triple shield could capture disinformation, misinformation, and malinformation, so long as Category 1 regulated services include such 'relevant content' within their terms of service. In doing so, it could create two issues.

Firstly, it only applies to Category 1 user-to-user services. Depending on where the cut off between Category 1 and 2B services is drawn and how Ofcom categorises certain services (it will be determined by user numbers, independent of functionality, and factors deemed relevant by the Secretary of State¹⁵⁸) means we could be faced with a situation where services with relatively large user numbers are not subject to the duty, yet smaller, 'high-harm', platforms are, which could create disparity among services in how content is treated. This becomes more concerning when one considers that the threshold conditions for differentiating between the categories will be determined by a statutory instrument. As stated above, this potentially exposes the categorisation process to pressure and lobbying from services. It is conceivable that services will use the resources at their disposal to influence the threshold conditions, so they fall within Category 2B. This is worrying from a public sphere perspective when you bear in mind: (i) that potential lobbying from services to influence the threshold conditions will not play out in a publicly accessible forum (as is the case with parliamentary debates on primary legislation), so the public will be largely unaware of this influence, and its impact on the shape of the regime; (ii) the social purpose of

¹⁵⁴Consequently, Harbinja argues in the report, and has argued previously, for the need to introduce a standard of 'manifest illegality' instead of the 'reasonable grounds to infer'. E Harbinja and N Ni Loideain, *Policy Report: Making Digital Streets Safe? Progress on the Online Safety Bill*, June 2023, IALS and Aston University.

¹⁵⁵OSA, s71(1), 72(3)(a).

¹⁵⁶*ibid*, s72(5).

¹⁵⁷*ibid*, s74(5).

¹⁵⁸See (n 118–123).

free speech, as it is likely that most users will be unaware of whether a service is under a duty or not, and therefore the extent to which it moderates the information available to them through the service; and (iii) that where users are aware that a service is under a duty this could result in some of those users moving to smaller, and in some cases more polarised, platforms, such as Truth Social and Rumble. This could lead to marginalisation of certain groups and could contribute to a more polarised public sphere – which as discussed above, is the opposite of what Mill argued for in his argument from truth. In doing so, it not only conflicts with the spirit of Mill's argument, and the marketplace of ideas theory, but it undermines the social purpose of free speech, embedded in these theories and in the argument from public discussion.

Secondly, by placing the impetus on regulated services to deal with false information through their own terms of service, this is another example of the OSA reneging on the government's intention to do-away with self-regulation, potentially placing self-regulation on a statutory footing. The problem with this from a public sphere and free speech perspective is that leaving it up to services to decide what content is covered in their terms of service and relying on them to apply those terms of service consistently puts decisions on our freedom of speech in the hands of the likes of Facebook and X. This is concerning when you consider the nature of terms of service, and users' relationship with them: they are often complex, and difficult for users to interpret and engage with, and they may be changed by the service at any point.¹⁵⁹ Although the OSA imposes a duty on services to include 'clear and accessible' provisions¹⁶⁰ in their terms of service, how this will be interpreted by services, and what this will look like in reality remains to be seen. Notwithstanding this, users' previous experience with terms of service generally may discourage them from engaging with OSA-compliant terms of services regardless of how 'clear and accessible' they are. Furthermore, under the OSA terms of service duties, it seems that users will have *a role* to play in identifying and notifying services of content that breaches the terms of service.¹⁶¹ The extent to which users will influence the respective service's application of the duty, and the consistency it is applied, remains to be seen. This is, perhaps, exactly what the proponents of removing the legal but harmful provisions were trying to avoid. This aspect of the shield provides a back-door for services to remove content that may be harmful but legal, and in doing so it undermines the social purpose of free speech, and in particular the argument from public discussion, as it means that Category 1 services will have the ultimate say in

¹⁵⁹Bernal, *Warts and All* (n 15) 127; S Zuboff, *The Age of Surveillance Capitalism* (Profile Books 2019) 48–50, 217–20.

¹⁶⁰OSA, ss10(8), 12(13), 15(7), 72.

¹⁶¹*ibid*, s72(5).

what we see or read, rather than us, as citizens, making an informed decision on what we are and are not exposed to. Moreover, this removal of content will happen behind closed doors. The process will, therefore, lack transparency and objective oversight.

'Recognised news publishers' and false information

The press's role in the dissemination of online false information relates to a point I made at the beginning of this article and is an issue in respect of the OSA has been hugely controversial. The regime provides exemptions for 'recognised news publishers'.¹⁶² 'News publisher content' posted on Category 1 services does not fall within the scope of the regime.¹⁶³ And section 180(1) exempts such publishers from the false communications offence. Pursuant to section 56(2) a recognised news publisher can be an entity that publishes news-related material, that is created by different persons, and is subject to editorial control¹⁶⁴ and, inter alia, which publishes such material in the course of a business, is subject to a standards code, has a registered office or other business address in the UK.¹⁶⁵ News-related material is defined at section 56(6) as material consisting of, inter alia, 'gossip about celebrities, other public figures or other persons in the news'.¹⁶⁶ Thus, the OSA arguably provides an exemption for large swathes of our press and media to publish content that is very often, and largely based, on misinformation and, at times, disinformation – which goes to the core of the trade-in celebrity gossip. In doing this, rather than protecting us from false information, the OSA could contribute to the distortion of the public sphere, as these false stories may (and sometimes do) become the dominant view,¹⁶⁷ thereby perpetuating the flaws advanced above in relation to the theories underpinning the ECtHR's free speech jurisprudence that should, paradoxically, serve to justify the OSA's existence.

Of course, such content may be covered by the torts of misuse of private information (MPI), or defamation, and could amount to breaches of the IPSO Editors' Code of Practice¹⁶⁸ and the Impress Standards Code.¹⁶⁹ But, in reality, these legal and regulatory mechanisms for protecting individuals and organisations from the publication of such content, or vindicating them from the damage that flows from it, are often ignored by the factions

¹⁶²*ibid*, defined in section 56. See also: Department for Culture, Media and Sport, Guidance: *Fact sheet on enhanced protections for journalism within the Online Safety Bill*, 23rd August 2022.

¹⁶³OSA, s18.

¹⁶⁴*ibid*, s56(2)(a)(i), (ii).

¹⁶⁵*ibid*, (a)-(g).

¹⁶⁶OSA, s56(6)(c).

¹⁶⁷Coe (n 6) 82–83.

¹⁶⁸See Clause 1 (Accuracy): <https://www.ipso.co.uk/editors-code-of-practice/>.

¹⁶⁹See Code 1 (Accuracy): <https://www.impressorg.com/standards/impress-standards-code/our-standards-code/>.

of the press that routinely publish this type of content.¹⁷⁰ This is because the revenue generated from publication outweighs the cost of any litigation and damages that *may follow* as a result. Indeed, because of the costs involved in bringing MPI or defamation claims, their use tends to be limited to individuals and organisations with the financial resources available to pursue these claims, yet, as Lord Justice Leveson acknowledged in his *Inquiry into the Culture, Practices and Ethics of the Press* the publication of this type of content causes ‘real harm to real people’.¹⁷¹ By this, he meant that press abuses not only affect a small minority, such as celebrities, sports men and women, or politicians, but also ordinary people. These people do not tend to have the financial resources required to fund litigation, nor do they have access to lawyers or reputation management and public relations advisors to help them respond to and spin stories based on false information.¹⁷² Consequently, it is ‘ordinary’ people who, because they have piqued the interest of the press and are therefore, for a limited time, ‘other persons in the news’¹⁷³ who stand to be exposed and caused the most damage by section 56(6), yet it is these ‘ordinary’ people who need the most protection.

Scratching beneath the surface there is a further issue here. By stipulating that recognised news publishers, and therefore those who are exempt, must produce news-related material that is ‘created by different persons’, is ‘published in the course of a business’, and has a ‘registered office or other business address in the UK’, the OSA potentially excludes many independent news publishers from the exemption because, many work remotely, and can be based abroad (while serving a UK audience), and/or they are run by a single person, which is often the case for hyperlocal publishers. And, in the case of many citizen journalists, they are not necessarily operating in the course of a business.¹⁷⁴ Whilst, at the same time, these criteria could create a loophole that could be exploited by those who want to avoid accountability, whether that be a news publisher engaging in the dissemination of false information, or a group of individuals masquerading as a news publisher who will use that position to plant false information (or other harmful content) – all they need to do is simply write a standards code, set up a complaints process, and call themselves a news publisher.¹⁷⁵

¹⁷⁰For further analysis of this issue, see: P Coe, ‘Press Regulation in the United Kingdom in a Changed Media Ecosystem’ in P Wragg and A Koltay (eds) *Global Perspectives on Press Regulation* (Hart 2023) 209–34.

¹⁷¹Lord Justice Leveson, *An Inquiry into the Culture, Practices and Ethics of the Press: Report* (HC 780, 2012) 50, [2.2].

¹⁷²Wragg (n 18) 60–61.

¹⁷³*ibid.*

¹⁷⁴Coe (n 6) 269–70; Independent Media Association, Response to the Online Safety Bill, 28th February 2023.

¹⁷⁵Press Recognition Panel, <<https://pressrecognitionpanel.org.uk/why-is-amendment-126-of-clause-50-of-the-online-safety-bill-so-important/>>; N Sparkes, Hacked Off analysis: Russell Brand’s Rumble channel may benefit from press loophole in Online Safety Bill, 2nd October 2023: <<https://>

Although, as I suggest in the conclusion to this article, I do not think that regulation is a panacea a solution would be for ‘news publishers’ to be defined in law as members of the press recognition system-approved regulator, which is currently Impress.¹⁷⁶ On the face of it, this would provide a clear definition, and remove the potential for malicious exploitation. However, bearing in mind much of the legacy press’s resistance to approved regulation since its creation post-Leveson,¹⁷⁷ this solution would likely fall at the first hurdle.

What this means for independent news publishers and citizen journalists is that their content is treated in the same way as other non-news material content when a conflict arises between content and the OSA’s safety duties – meaning their content is significantly under-protected in comparison to exempt ‘recognised news publishers’ content. As I have argued in a previous article, albeit in the context of hate speech,¹⁷⁸ and revisit briefly here, this is because, sections 22 and 33 set out a general duty applicable to user-to-user and search services respectively to ‘have particular regard to the importance of: (i) ‘protecting users’ right to freedom of expression’ and (ii) ‘protecting users from a breach of any statutory provision or rule of law concerning privacy’. Additionally, section 17 provides ‘duties to protect content of democratic importance’ and section 19 prescribes ‘duties to protect journalistic content’. Unlike the sections 22 and 33 duty, the sections 17 and 19 duties only apply to ‘Category 1 services’. The fact that the core free speech duties pursuant to sections 22, and 17 and 19 of the Act only require services to, in respect of section 22, ‘have particular regard to’ the importance of freedom of expression, or in the case of sections 17 and 19 ‘take into account’, free speech rights or the protection of democratic or journalistic content, means that regulated services may simply pay lip service to these ‘softer’ duties when a conflict arises with the legislation’s numerous and ‘harder-edged’ safety duties.

In respect of the section 22 duty, 22(4)–(7) requires Category 1 services to carry out and publish impact assessments of their safety measures and policies on users’ freedom of expression. However, regardless of this, the distinction between the harder and softer duties could encourage services to publish assessments with boiler plate answers. Similarly, for sections 17 and 19, services may produce template policies that say those services have ‘taken into account’ the protection of democratic or journalistic content. So long as they can point to a small number of decisions where moderators have had regard to, or taken these duties into account, they will be able to

hackinginquiry.org/russell-brands-rumble-channel-may-benefit-from-press-loophole-in-online-safety-bill/> accessed 7th January 2024.

¹⁷⁶For an overview of the system, see: Coe (n 70) 209–34.

¹⁷⁷*ibid*; Wragg (n 18).

¹⁷⁸Coe (n 2).

demonstrate their compliance with the duties imposed by the OSA to Ofcom. It may be extremely difficult, or perhaps even impossible, to interrogate the process.¹⁷⁹

This situation is concerning for free speech when you consider that many of these independent journalists are increasingly stepping into the watchdog shoes of the press, by making valuable public interest contributions to our public sphere¹⁸⁰; whereas there are a number of exempt entities that profit from publishing celebrity gossip and, incidentally, false information.¹⁸¹ Therefore, inadvertently, in this regard, the OSA could contribute to the marginalisation, not only of these types of publishers, but also the individuals and communities they give a platform to, including minority and under-represented groups¹⁸² and, in doing so, it could disproportionately impact on certain types of reportage and information. This limits the scope of the public sphere, and the breadth and depth of public discourse and, in turn, in conflict with the principles underpinning the ECtHR's Article 10 jurisprudence discussed above, it could restrict equal opportunities for citizens to engage in public discussion and contribute to the governance of their communities.

In conclusion, the proliferation of online false information highlights the flaws in the theories underpinning the ECtHR's free speech jurisprudence, thereby justifying the OSA provisions considered above. Yet, paradoxically, for the reasons advanced throughout this article, the regime conflicts with, and is disconnected from, the Court's jurisprudence, and the spirit of its theoretical foundations, and therefore has the potential to interfere perniciously with the right to free speech. In essence, this is because, by making services responsible for the content on their platforms, the OSA enhances their ability to undertake a behind-the-scenes, and what may often be an invisible, curating role. Although this is not new – with privatised censorship always existing online – the OSA's triple shield gives regulated services a statutory basis for subjectively evaluating and censoring content, by providing a backdoor for them to remove content that may be harmful but legal. As I have previously suggested, this, along with the potential conflict between the harder and softer duties, could lead to platforms adopting an over-cautious approach to monitoring content by removing anything that *may* bring them within the scope of the duties and regulatory sanctions.¹⁸³ This risk is amplified by the lack of clarity around the false communications offence that I have highlighted, which could lead to legitimate content being removed

¹⁷⁹Coe (n 2) 68–69.

¹⁸⁰See generally: Coe (n 6).

¹⁸¹See note 18; *The Cairncross Review* recognised that to increase online advertising revenue, newspapers have encouraged the sensationalisation of news and the prioritisation of low-quality 'clickbait' over high-quality, investigative and minority publications: *Cairncross* (n 27) 42–44.

¹⁸²Coe (n 6) 90.

¹⁸³Coe (n 2) 70.

because it is incorrectly thought to be illegal and/or merely harmful on the basis of falsity. And, cynically, for these reasons, the OSA may provide services with an opportunity, or an excuse, to remove content that does not conform with their ideological values on the basis that it could be illegal or that it breaches its terms of service.¹⁸⁴

Conclusion

Can we solve the problem of online false information through regulation alone? Unfortunately, the short answer to this question, and whether there is *a solution* to dealing with online false information is no – there is no silver bullet. It has existed for centuries, and always will exist, and so long as we have the internet and social media, it will be pervasive and invasive. I am not convinced that the OSA, and regulation generally, is the appropriate way to meet the challenge we face, as it is not the radical panacea for protecting us from online harms that it has been presented as. Indeed, history tells us that law and regulation to deal with false information, in whatever form it comes in, tends not to work, largely because there is often a tension between misinformation in particular and free speech laws and principles¹⁸⁵ – which I have highlighted throughout this article. In my view, when it comes to something as amorphous as false information, the rigidity of legal regulation means it is, probably, doomed to fail – and may, in some cases, make things worse. With the OSA, I fear, for the reasons I have explained in the preceding sections, that despite the important reasons for its enactment, and the commendable principles upon which it is based, by taking a regulatory step forward in tackling online harm, the duties the Act places on services, and the incidental power it gives them, could have unintended and insidious implications for free speech.

Rather, in addition to ensuring that Ofcom, the Information Commissioner's Office, and the press regulators are properly funded, resourced, and supported, we must find other ways to live with false information without inadvertently eroding our right to free speech. There is no one way of doing this. But there are imperfect (and I acknowledge, somewhat idealistic) 'more speech'¹⁸⁶ solutions that can be deployed to improve the situation, and meet the challenges posed by false information, that accord with the ECtHR's case law, and the free speech theories which underpin it.

Firstly, and fundamentally, we must take responsibility for our own online behaviour, and in doing so acknowledge that behavioural adaptation may be required at a (micro) individual and (macro) societal level. For instance, we

¹⁸⁴ibid; Cram (n 16) 133–34.

¹⁸⁵Bernal, *The Internet, Warts and All* (n 16) 248–50.

¹⁸⁶Cram (n 16) 134.

must improve our education and awareness on how to lead a safe and productive online life (to be aware of the ‘dangers’, in its broadest sense, of co-existing online and offline, but also to be able to take advantage of its many benefits). This should start from primary school, with media or digital literacy forming a fundamental part of our national curriculum. Within this, children should be taught, from as young as possible, how to be safe online, which would include learning about false information and how to identify it, and then ‘manage’ it.¹⁸⁷ It is important for me to acknowledge at this point that I am not making a new or revolutionary call to action, at least in substance. Indeed, Ofcom is under a duty, pursuant to section 11 of the Communications Act 2003, to promote media literacy, including to build citizens’ resilience to false information.¹⁸⁸ Additionally, in 2017, the Children’s Commissioner’s report, *Growing up Digital*, called for the creation of a compulsory digital citizenship programme for pupils aged 4–14, to improve children’s digital literacy skills and digital resilience, and to broaden digital literacy education beyond safety messages.¹⁸⁹ This was followed, in 2018, by the House of Lords Select Committee on Political Polling and Digital Media, which stressed the need to teach critical literacy skills in schools, to limit the spread of misinformation online and its potential impact on democratic debate.¹⁹⁰ In February 2019, the House of Commons Digital, Culture, Media and Sport Select Committee, in its report *Disinformation and ‘Fake News’: Final Report*, called for digital literacy to be the fourth pillar of education, alongside reading, writing and mathematics.¹⁹¹ Subsequently, in June 2019, the UK government’s Department for Education published its ‘Teaching online safety in schools’ guidance.¹⁹² Yet despite all of this activity, in November 2020 it was announced that an All-Party Parliamentary Group on Media Literacy intended to commission an independent inquiry into media literacy in schools, led by the former chair of the DCMS, Damian Collins MP.¹⁹³ This was because research had found that only half of teachers had heard

¹⁸⁷P. Coe, ‘Freedom of Speech and the Regulation of Fake News in the United Kingdom: Managing Misinformation and Disinformation, and Protecting Free Speech, in the UK’s Modern Media Ecology’ in O. Pollicino (ed) *Freedom of Speech and the Regulation of Fake News, Ius Comparatum – Global Studies in Comparative Law* (Intersentia 2023) 503–39, 527–28.

¹⁸⁸Ofcom, ‘Making Sense of Media’: <<https://www.ofcom.org.uk/research-and-data/media-literacy-research>> accessed 13 December 2023.

¹⁸⁹Children’s Commissioner, *Growing Up Digital: A Report of the Growing Up Digital Taskforce* (January 2017) 3.

¹⁹⁰House of Lords Select Committee on Political Polling and Digital Media, *Report of Session 2017–19* (HL Paper 106, 17 April 2018) [319].

¹⁹¹House of Commons, Digital, Culture, Media and Sport Committee, *Disinformation and ‘fake news’: Final Report* (HC 1791, 14 February 2019) [308], 86.

¹⁹²See UK government, Department for Education, *Guidance Teaching Online Safety in Schools*, 12 January 2023.

¹⁹³The APPG, convened by media literacy charity The Student View, includes MPs and peers from the Labour, Conservative and Scottish National Parties: <<https://www.parliament.co.uk/APPG/media-literacy>> accessed 14 December 2023.

of the government's guidance, with only 14 per cent of schools implementing its recommendations.¹⁹⁴ Thus, what I am suggesting here is that this call needs to be *renewed*, with greater impetus and urgency, ensuring that Ofcom, for instance, is properly funded and resourced to meet its statutory obligations. And, importantly, media and digital literacy should not just start and finish with children. It should be available and accessible to all citizens regardless of age. It should form part of teacher training, and perhaps even be integrated into further and higher education courses.

Secondly, we must engage more directly and assertively with organisations that offer services, resources, and expertise to tackle false information. This requires cross-party commitment from our political parties to appropriately empower and fund these organisations to ensure they are able to implement long-term strategies. Thirdly, we must utilise our research power through our universities and other research and funding bodies to continue to develop our understanding of the problem, and ways we can tackle it. Again, this requires a cross-party commitment to long-term funding.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributor

Peter Coe is an Associate Professor in Law at Birmingham Law School, University of Birmingham. He is a Research Fellow at the Institute of Advanced Study, Durham University, a Senior Visiting Research Fellow at the School of Law, University of Reading, and an Associate Research Fellow at the Institute of Advanced Legal Studies and Information Law and Policy Centre, University of London.

¹⁹⁴MPs Form Group to Safeguard Children from Fake News', *Society of Editors*, 25 November 2020.