Routledge
Taylor & Francis Group

Journal of Media Law

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rjml20

Online harm, free speech, and the 'legal but harmful' debate: an interest-based approach

Konstantinos Kalliris

To cite this article: Konstantinos Kalliris (2024) Online harm, free speech, and the 'legal but harmful' debate: an interest-based approach, Journal of Media Law, 16:2, 390-416, DOI: 10.1080/17577632.2024.2425547

To link to this article: https://doi.org/10.1080/17577632.2024.2425547

Published online: 02 Dec 2024.

JOURNAL OF MEDIA LAW
2024, VOL. 16, NO. 2, 390–416
https://doi.org/10.1080/17577632.2024.2425547

Routledge
Taylor & Francis Group

Online harm, free speech, and the 'legal but harmful' debate: an interest-based approach

Konstantinos Kalliris

Essex Law School, University of Essex, Colchester, UK

ABSTRACT

The Online Safety Act introduced to public and academic debate the concept of harmful but legal online content. The idea that non-wrongful speech ought to be regulated in certain circumstances, is prima facie controversial, as it appears to be incompatible with the Harm Principle, which remains one of the most common considerations for legislators in liberal democracies. Most arguments against this provision have focused on the right of speakers to express themselves freely provided that they do not wrongfully harm others. This paper argues that, while some of these concerns for the rights of speakers are legitimate, it is listeners who would see their right to free speech mostly affected by the regulation of legal but harmful online speech, due to its distinct paternalistic implications. While the idea behind the provision is justifiable, the means employed and the persons or organisations entrusted with the task are crucial from a normative standpoint.

ARTICLE HISTORY Received 29 February 2024; Accepted 31 October 2024

KEYWORDS Online safety act; freedom of speech; online harm; paternalism

Introduction

The Online Safety Act 2023 undertakes a complex but important task: the protection of internet users from online threats. In this context, the concept of 'legal but harmful' content was introduced: internet service providers would be required to regulate content that is not illegal but can cause harm to others. This provision was heavily criticised by various stakeholders mostly on the grounds that it paves the way to unjustifiable restrictions of free speech. This would not be without precedent: regulation of free speech is notoriously problematic in most contexts. What makes the notion of regulating legal but harmful speech on the internet worthy of special attention is its focus on offence and the means available to the

CONTACT Konstantinos Kalliris konstantinos.kalliris@essex.ac.uk

regulators. Restrictions of free speech that may result in physical harm, such as incitement to violence, are relatively uncontroversial (mostly due to their perceived conformity with the Harm Principle), but regulating offensive speech is more difficult to justify. We should not, of course, infer from this difficulty that offensive speech is necessarily beyond regulation: on the contrary, it is important to acknowledge, both morally and legally, that speech can cause psychological harm and that psychological harm is real harm. In this spirit, the Online Safety Act set out to protect internet users from the psychologically adverse effects of expressions (and, inevitably, information) that may be offensive or distressing. In doing so, it would have to grapple with a very elusive concept; entrust internet providers with enforcing the necessary restrictions; identify appropriate means; and assume that valuable or merely useful information would not become inaccessible (at least online, as it could still be otherwise available). It is reflective of the complexity of this largely unprecedented task that the proposal did not survive public scrutiny.

The identification of appropriate means is a well-known puzzle in theoretical discussions of autonomy-restricting legislation. The reason is that means matter in a normative sense. Prison is a very good example: the acknowledgment that one deserves to have one's autonomy restricted as punishment for a wrong committed does not necessarily entail that imprisonment is justified as well. This is because imprisonment has devastating and indiscriminate effects on autonomy. It is not, then, enough to determine whether autonomy-threatening or rights-limiting regulation is justified in principle – the appropriate means must also be identified. In the context of our present enquiry, the question of means is even more pressing because the online world offers opportunities which are often unavailable to 'real world' regulators. Users guilty of legal but harmful speech could be banned or have their posts removed. This would resemble real-life coercion or compulsion and is normally reserved for illegal speech. But the potential victims of harmful speech can be protected in other ways: for example, questionable posts can be hidden behind a warning or 'downgraded' so that they do not appear on one's timeline as easily. These subtler means are often presented as much less controversial and mostly respectful of autonomy, because users can still post their content or choose to access the regulated speech. However, even subtler ways of effectively removing options (by making them difficult to find or labelling them as potentially harmful) can have adverse effects on the exercise of autonomy, *especially for the potential receivers of the regulated speech*. This largely neglected point will be an integral part of the argument presented here.

The 'legal but harmful' provision did not survive, but, as these introductory comments suggest, the issue remains pressing: given the extent of online conduct that can harm the well-being of users (online bullying and

misinformation are obvious examples), is it a good idea to regulate otherwise legal speech and, if so, what are the principles that ought to guide such legal provisions? These questions are primarily normative: we cannot answer them unless we grapple with the right to free speech and why we consider it valuable. The best way to approach the matter is by looking at the right in the context of a general theory of rights: why do we value rights in general? The answer to this question will help us understand potential conflicts between freedom of speech and other rights or values. Finally, we will look at the role of different 'players' in the game that is the online world: there are users who mostly transmit information and users who mostly receive it; there are 'referees' (the state or online service providers); there are groups with special interests; and so on. This mostly theoretical discussion will allow us to have a clearer picture of what is really at stake when legal but harmful online content is regulated. This article will attempt to do this work through the lens of an interest-based account of the right to free speech. I will first defend an account of rights in general and of the right to free speech in particular that focuses on interests and is at least appropriate for current purposes; then, I will provide a sketch of a -hopefully- uncontroversial conception of paternalism; finally, I will argue that the regulation of legal but harmful content for adults is unjustifiably paternalistic predominantly for receivers of free speech (listeners) rather than for the speakers themselves.

Rights, interests, and freedom of speech

Why do we need rights and why do we cherish them? The obvious answer is that they give us power over others: in the oft-quoted Millian terminology, they give us a *claim* on others to act in a certain way.[1] This is an intuitive answer, but it clearly requires further unpacking, especially since the mere existence of a claim does not necessarily entail the existence of a right (at the very least it must be a valid claim).[2] There are, therefore, certain claims we are entitled to make on others that correspond to rights and, consequently, carry special normative weight. How can we tell where this weight comes from? The question remains pressing even if not all rights correspond to duties in the Hohfeldian sense,[3] being more accurately classified as powers or immunities. In the relevant literature, two families of theories dominated most attempts to answer this question: will theories and interest theories. While it is beyond the scope of this work to contribute to a debate that has been raging for decades, I must briefly explain why continuing this

[1]JS Mill, *Utilitarianism*, G Sher (ed), (Hackett 2002) 54.
[2]Joel Feinberg, 'The Nature and Value of Rights' (1970) 4 Journal of Value Inquiry 243, 257.
[3]Wesley Newcomb Hohfeld, 'Some Fundamental Legal Conceptions as Applied in Judicial Reasoning' (1913) 23(1) Yale Law Journal 16.

discussion on the assumption that interest theories are at least partly defensible does not hurt my argument or require an unconditional acceptance of any particular interest theory. I will not discuss demand theories separately, because I take them to make similar assumptions to will theories in the sense that they also focus on human agency as the source of a particular kind of demand that the right-holder can make of others.[4]

Both families of theories have intuitive appeal. Will theories conjure up images of the autonomous person who is both able and entitled to control her life and how others ought to treat her. In Hart's famous words, to hold a right is to have a kind of *small-scale sovereignty*.[5] Being a sovereign in this sense means that one has the power to make a valid claim that others act in a specific manner, as well as the power to waive one's rights. For example, my right to free speech would entail that I have a claim against censorship which I can direct at others (the duty-bearers). The right would also entail the power to consent to censorship. The normative appeal of will theories is obvious: if they are defensible, we are guaranteed to control our personal sphere and what others can or ought to do for us. Crucially, this power to control stems from the ability to do so, which in turn reflects the necessary connection between choice and right. A necessary and, for many, fatal consequence of this connection is that we end up with far less right-holders than desirable. The mentally challenged, patients in a comma and even newborn babies generally have little to no ability to make choices and exercise the control required for will theories to make sense, especially if we understand rights as contributing to autonomy and self-realisation. Even if the will theorist can find a way around this by defending, for example, the view that a general ability (including past and future ability) for such control and sovereignty is enough, regardless of whether it is presently possible, the rights of some humans and most animals would be impossible to accommodate. Yet, many people would think that it should be clear and unquestionable that all these beings ought to have at least some rights. There seems to be no room in will theories for children or animals to have rights in the strict sense – that is, rights that make them 'small scale sovereigns' rather than rights exercised by others on their behalf.[6]

Furthermore, and more importantly for this discussion, will theories share another characteristic that many find indefensible: by granting individuals this level of control over their small-scale sovereignty (and by making this control essential to having rights), the will theorist is bound to accept that

[4]John Skorupski, *The Domain of Reasons* (Oxford University Press 2010) 310–311.
[5]HLA Hart, *Essays on Bentham: Studies in Jurisprudence and Political Theory* (Clarendon Press 1982) 183.
[6]MacCormick makes the point very convincingly regarding the rights of children in Neil MacCormick, *Legal Right and Social Democracy: Essays in Legal and Political Philosophy* (Oxford University Press 1984) 155–157.

all rights can be waived. It is, in other words, for the right-holder to decide whether she will exercise her rights, seek the protection of the state, or allow their full breach. This appears sensible and indeed unquestionable as far as certain rights are concerned: I should, for example, be allowed to demand the protection of my personal property by law or allow others to use it, even permanently. However, there are rights which appear to be unwaivable: should we be free to choose torture or slavery? Even Mill, arguably the most influential proponent of free choice and anti-paternalism, thought that any contract that would result in the enslavement of a human being should be 'null and void; neither enforced by law nor by opinion'.[7] In fact, these rights are among the most important ones and any theory that makes them dependent upon choice or contract is at a serious disadvantage. I think that the point can be made that some aspects of the right to free speech are, at least indirectly, akin to an unwaivable right. I will say more about this later.

Interest theories of rights are equally intuitive, as they associate rights with personal well-being. They are also capable of accommodating unwaivable rights, as well as the rights of beings with limited cognitive capacities, including animals.[8] The main weakness of interest theories is that, clearly, not all interests produce rights: it is in my interest to win a race, but I have no right to win it and I certainly have no claim against others to let me win. Even if this win is important for my well-being (because, say, the prize is a significant amount of money and I am starving), no one can be said to be under duty to let me win. Therefore, it would be more accurate to say that sometimes one's personal well-being is sufficient reason to recognise that one or more persons have a certain duty towards the right-holder.[9] This approach clarifies that the existence of an interest is a necessary but not a sufficient condition for the existence of a right. However, what exactly counts as an interest that is sufficiently important for one's well-being to create a right is a question with no clear answer. At the same time, there seem to be universally accepted rights that do not stem from the interests of the right-holder. The most famous example in the literature is the right of a journalist not to reveal her sources. Joseph Raz argues that, in this case, it is rather the interest of the public to an independent media that provides the grounds for acknowledging this right. The same logic applies to the right of lawyers and doctors to maintain confidentiality: it is in the interest of the community to have these rights in place.[10] But as Frances Kamm points

[7] J S Mill, *On Liberty* (Dover Publications 2002) 86. See also Neil MacCormick, 'Rights in Legislation', in P Hacker and J Raz, (eds), *Law, Morality and Society: Essays in Honour of HLA Hart* (Oxford University Press 1977) 197.
[8] Leif Wenar, 'The Nature of Rights' (2005) 33(3) Philosophy and Public Affairs 223, 241.
[9] Joseph Raz, *The Morality of Freedom* (Oxford University Press 1986) 166.
[10] Raz (n 9) 179.

out, this justification can only mean that the journalist's (or the lawyer's or the doctor's) interests are *not* sufficient to give rise to these rights.[11] All rights associated with specific roles or offices seem to require this appeal to the common good, an appeal that clearly departs from the individualistic character that most people assign to rights. This response seems to summon the worst enemy of any robust theory of rights: utilitarianism. It is also a reminder of the main advantage of will theory: as long as we enjoy small-scale sovereignty, we secure the ability to make suboptimal or even bad decisions.

In certain versions of utilitarianism, the morally justified arrangement is the one that *normally* produces more utility rather than the one that actually does so in each scenario. Interest theories of rights can take a similar path: to acknowledge a certain right to a certain group of individuals is to presuppose that the exercise of this right is, *in all normal circumstances*, good for them.[12] This approach reminds us why right-holders are not expected to always exercise their rights in the pursuit of the good (not only because that would be impossible and absurd but also because one's circumstances are not always normal). At the same time, it explains why it is necessary to maintain some unwaivable rights, contrary to traditional will theories. Rights are indeed strongly related to freedom: they give us the necessary space and options to lead an autonomous life. It is plausible to think that, at least to an extent, the exercise of autonomous choice, as protected by rights, draws its value from its capacity to secure a good life.[13] Will theories of rights seem to be particularly capable of safeguarding the space required for autonomy and self-authorship (in Isaiah Berlin's famous terminology,[14] this would amount to negative freedom), but not as accommodating as far as our ability to pursue options (positive freedom) is concerned. Unwaivable rights serve to maintain a basic degree of this ability at all times, even if the right-holder does little to use it in a meaningful manner. For those who think that one of the most fundamental duties of government is to protect the welfare of its citizens, this account is a valuable guide in the description and interpretation of legal rights.

We are still left with one problem: what about the 'rights' of role/office holders? I think that the best response to this objection is that these are not rights in the first place. In the traditional Hohfeldian sense,[15] the journalist seems to have the *liberty* not to disclose her sources, in the sense that she is not under duty to do so. And the judge who appears to have the right to

[11]FM Kamm, 'Rights' in J. Coleman and S. Shapiro (eds), *The Oxford Handbook of Jurisprudence and Philosophy of Law* (Oxford University Press 2002) 485.
[12]MacCormick (n 6) 163.
[13]Konstantinos Kalliris, 'Self-Authorship, Well-being and Paternalism' (2017) 8(1) Jurisprudence An International Journal of Legal and Political Thought 23, 31; Raz (n 8) 281. Raz, of course, makes the stronger statement that autonomy is only valuable when exercised in pursuit of the good.
[14]Isaiah Berlin, *Four Essays on Liberty* (Oxford University Press 2017).
[15]Hohfeld (n 3).

sentence (a right that ought not to stem from any interest of hers) has the *power*, rather than the right, to send someone to prison.[16] As far as claim-rights are concerned (that is rights that are enforceable, even by coercive means), interest theory seems to cover a lot of ground. For current purposes, we need nothing more than this acknowledgement. Even if interest theory cannot account for everything we refer to when we speak of rights, even if we need a hybrid theory[17] to cover all those instances, the idea that our claim-rights normally correspond to an interest that is grounded on personal well-being is enough for the next steps of my argument. In the next section, I will explain why it is the most accurate way to describe the value of free speech, especially in its manifestation as a legal right.

The right to free speech

Freedom of speech has been defended from various viewpoints. Some, like Thomas Scanlon, believe that it is a constraint on government action, since suppressing or regulating free speech would not treat citizens as autonomous beings capable of critical thinking.[18] Others maintain that free speech is essential for democracy[19] and that, without it, the actions of a democratic government lack legitimacy.[20] Experience seems to support this view: it is precisely in a democratic state that the law is commonly called upon to describe the legal right to free speech and determine its limits. Experience also teaches us that, in doing so, the law can go wrong in two different directions. First, it can protect less free speech than desired, by banning expressions that ought to be allowed. Relevant examples include blasphemy laws, laws that ban expressions considered unpatriotic (such as burning the flag) or laws that supress criticism against public officials. Second, the law can end up allowing too much free speech. Hate speech is an obvious example of how legislators can fail in this respect, but it is not necessarily the only one: consider grieving families which are subjected to listening to the offensive chants of a crowd protesting against the actions of their relatives while they were alive. Apart from its direct effects on democracy and the well-being of individuals (as in the familiar example of someone yelling 'fire' in a full theatre), speech seems capable of causing extreme discomfort that may amount to psychological harm, and it is the law's business to maintain the very delicate balance between

[16]Matthew Kramer, 'Rights Without Trimmings' in Matthew Kramer, N E, Simmonds and Hillel Steiner (eds), *A Debate Over Rights: Philosophical Enquiries* (Oxford University Press 2000) 9.
[17]For such an attempt see Gopal Sreenivasan, 'A Hybrid Theory of Claim-Rights' (2005) 25 Oxford Journal of Legal Studies 257.
[18]Thomas Scanlon, 'A Theory of Freedom of Expression' (1972) 1(2) Philosophy & Public Affairs 204.
[19]Joshua Cohen, 'Freedom of Expression' (1993) 22(3) Philosophy & Public Affairs 207.
[20]Joshua Cohen, 'Deliberation and Democratic Legitimacy' in James Bohman and William Rehg (eds), *Deliberative Democracy: Essays on Reason and Politics* (MIT Press 1997) 67.

freedom of speech and protection from wrongful harm. It will be easier to appreciate -and, hopefully, untangle- the complexity of this task with a clearer view of what it is exactly that a right to free speech is meant to protect. It is impossible to do justice to the depth and sophistication of the relevant debate here. However, I will attempt to focus on specific aspects of some of the relevant theories, which I believe to be directly relevant to the issue of freedom of online expression.

The starting point of every attempt to understand the value of free speech and why a legal right to free speech is necessary is the acknowledgment that usually (but not necessarily) there are two parties involved: the speaker and the listener. We may call the right of the former the *active* right to free speech and that of the latter the *passive* right to free speech. The active right to free speech is usually defended on the grounds of what Joshua Cohen describes as the *expressive interest*: 'a direct interest in articulating thoughts, attitudes, and feelings on matters of personal or broader human concern, and perhaps through that articulation influencing the thought and conduct of others'.[21] The last part of Cohen's description is, as he acknowledges by the qualifier 'perhaps', controversial. First, not all speech amounts to communication, as we often express ourselves without intent to communicate anything[22] – this is why I said earlier that the existence of two parties is not necessary for the protection of free speech. Speech that is not meant to be received by anyone or, perhaps more accurately, speech that is addressed to the speaker herself, such as diaries and self-motivational speeches, cannot be reasonably excluded from any legal protection granted to speech that is clearly communicative.[23] Second, the view that one's interest in influencing the thought and conduct of others is sufficiently important to support a right to such influence is indefensible (in part, because others may have an interest not to be influenced). A better way to put it would be to include in this conception of the active right to free speech the *opportunity* to validate one's views and life choices through free expression, without that entailing that one has the right to access all available platforms or actually persuade others to follow.[24] This expressive interest, thus qualified, may not guarantee an audience for the speaker, but it requires legal protection from undue interference from the state or others. If the opportunity to express our views and validate our ways of life in our political communities is as important as I have suggested, restricting the active right to free speech because of its unpleasant or controversial content (rather than its direct harm to others) would be very difficult to justify.

[21]Cohen (n 18) 224.
[22]C Edwin Baker, *Human Liberty and Freedom of Speech* (Oxford University Press I989) 5I.
[23]Matthew Kramer, *Freedom of Expression as Self-Restraint* (Oxford University Press 2021) 23.
[24]Joseph Raz, *Ethics in the Public Domain: Essays in the Morality of Law and Politics* (Oxford University Press 1995) 156–157.

The passive right to free speech, on the other hand, is based on our interest to listen to what others have to say. While the most passionate defences of free speech in the public sphere normally focus on the active right to free speech, the passive right is much easier to describe and defend. People tend to listen more than they speak because it is in their interest to receive information. The passive right to free speech also extends beyond any narrow conception of communication, as it does not require the active participation of the speaker: many of the theorists referenced in this article will never discuss their work with me, but this has no effect on my undeniable interest in receiving their views. The flavour of this approach is distinctly Millian and, by necessity, consequentialist. Mill begins his discussion of free speech in the second chapter of *On Liberty* with the declaration that no government should 'determine what doctrines or what arguments [people] shall be allowed to hear'.[25] For Mill, all views have some use for the listener: if they are right one can adopt them; if they are wrong one can compare them to the truth and commit even more to the latter.[26] The idea that all views and arguments are useful to the listener is undoubtedly far-fetched. We know that the views of the Nazi regime were not only worthless but harmful. In fact, such manifestly false views cannot even claim to contribute to the development of our rational capacities.[27] The point is particularly important given that we can now be far less optimistic than Mill regarding the human capacity for rational deliberation.[28] In fact, false views are not the only ones that may harm this capacity: even an overload of otherwise useful information from a well-meaning source may result in poor rational deliberation.[29] This grim observation may have normative consequences not only for the passive right to free speech (since it is probably not in our interest to listen to *all* views), but presumably for active free speech as well: perhaps we should regulate what people say to protect others from the disutility of certain forms of expression or simply too much information.

What is then left for the right to free speech? One possible answer is that, as already mentioned, free speech is essential to democracy. The strongest defence of this view, famously articulated by John Rawls, is based on the idea that the former is constitutive of the latter: every time free speech is restricted, democracy is harmed. Rawls makes the point with reference to political speech in particular[30] and, therefore, flat earth theories, for

[25]Mill (n 7) 13.
[26]Ibid 14.
[27]David 0 Brink, 'Millian Principles, Freedom of Expression, and Hate Speech' (2001) 7(2) Legal Theory 119, 125.
[28]Brian Leiter, 'The Case Against Free Speech' (2016) 38 Sydney Law Review 407, 431.
[29]Emmanuel M Pothos et al, 'Information overload for (bounded) rational agents' (2021) 288 Proceedings of the Royal Society B 1.
[30]John Rawls, *Political Liberalism* (Columbia University Press 2005) 254.

example, would probably not enjoy the same level of protection. But the point can be made more generally: if Scanlon's view about the state's duty to treat its citizens as autonomous rational agents has any merit, all restrictions of free speech are potentially undemocratic. Friedrich Hayek extends the point to democratic self-government, arguing that without free speech or with free speech regulated by the governing elites, citizens will not have access to the vital information they need to make autonomous decisions.[31] It is no accident that philosophical anarchism's criticism of liberal democracy is based on the rejection of the view that such autonomous decisions are possible in the first place, given the very complicated knowledge one needs to access and assess to reach a conclusion in matters of, say, nuclear policy.[32]

Given that, pace Wolff's anarchism, democratic government is compatible with citizens not actively making autonomous decisions on all matters,[33] there is a less ambitious point to be made: if democracy is a way to secure self-government. and free speech is a necessary tool for democracy, the general assumption must be that citizens can, in normal circumstances, exercise their right to free speech meaningfully. And since most of us are both speakers and listeners, the assumption must be that we possess the reasoning ability both to assess the information we receive and to use it to express ourselves. If this were true, Hayek's connection between free speech and democracy would be easier to accept. But Wolff has a point: most of us can be poor listeners when it comes to many issues, including some far less complex than nuclear policy. And we can be equally poor speakers as well when what we wish to express goes beyond very basic needs and desires.[34] Not only is free speech not necessarily beneficial for democracy, but it can actively undermine it, as the antagonistic relationship between populism and democracy reveals.[35] Even if we resist the pessimistic view that we can articulate and understand very little that is truly useful for democracy, this is still a very thin defence of free speech: what about views that are manifestly false, worthless or antidemocratic? Even if they are not directly harmful for democracy, they are certainly not valuable in this sense, at the very least because they add to the noise and information overload mentioned above.

A better way to appreciate the value of free speech requires a holistic approach which transcends the distinctions mentioned so far. The active right to free speech, the passive right to free speech and the contribution of free speech to democracy are merely snapshots of what we truly value

[31]Friedrich Hayek, *The Road to Serfdom* (University of Chicago Press 1944).
[32]Robert Paul Wolff, *In Defense of Anarchism* (Harper Torchbooks 1970) 17.
[33]Harry Frankfurt, 'The Anarchism of Robert Paul Wolff' (1973) 1(4) Political Theory 405, 408.
[34]Leiter (n 28) 419–420.
[35]Koen Abts and Stefan Rummens, 'Populism versus Democracy' (2007) 55(2) Political Studies, 405.

about free expression. A more enlightening approach would see freedom of speech as engulfing two related and mutually dependent freedoms: freedom of communication and freedom of thought.[36] In her ground-breaking work, Seana Shiffrin argues that freedom of speech plays a 'special though not exclusive, role in the development of the mind and personality of each agent *qua* thinker'.[37] There is no doubt that 'thinkers' are essential for democracy: as Mill makes clear very early in *On Liberty*, the liberal political community he defends presupposes citizens 'capable of being improved by free and equal discussion'.[38] But this valuable effect of free speech is by no means the only one and human beings are thinkers for a number of other important reasons (including good reasons to express ourselves without communicating, for example by recording our thoughts in a journal). Despite the prevalence of arguments in support of free speech that emphasise the value of political speech, Shiffrin is right to point out that it is not at all obvious that, from a thinker's perspective, thoughts about, for example, mortality or friendship are less valuable.[39] In fact, people are very keen to maintain the right to hold their political views, but they are equally interested in shaping and sharing their opinions on morality, religion, human association and other areas of life which are directly associated with what we understand as our 'self'.

Once we appreciate this position, the connection between personal autonomy and freedom of speech becomes clearer: what we say and what we listen to is part of who we are and our effort to author our own lives. This self-authorship is an exercise of autonomy: it is not enough to mimic others or mindlessly follow their lead. To be a thinker in this sense, one needs both negative and positive freedom.[40] Freedom of speech in this sense includes the ability to refuse to express oneself in specific ways and this is not captured adequately by referring to the active or passive right to free speech, as Shiffrin correctly notes in her discussion of the mandatory allegiance to the flag in American schools. However, what is at stake here is not primarily the 'autonomous thought process of the compelled speaker'[41] as Shiffrin argues. What is predominantly at stake is her self-authorship: children, as well as adults, have a strong interest in being allowed to shape their basic beliefs unhindered by indoctrination or forced uncritical compliance that may be difficult to reject later in life. Again, this is easier to grasp if we focus on the right-holder's interests, since any brand of will theory cannot ignore the intellectual maturity of the children in question and their ability to make such

[36]Seana Valentine Shiffrin, *Speech Matters: On Lying, Morality, and the Law* (Princeton University Press 2016) 79.
[37]Shiffrin (n 36) 80.
[38]Mill (n 7) 8–9.
[39]Shiffrin (n 36) 93.
[40]Berlin (n 14).
[41]Shiffrin (n 36) 94.

choices. The right to free speech is important for self-authorship which in turn is crucial for personal well-being, at least in autonomy-supporting cultures.[42]

This brings us to my earlier point about the right to free speech as an unwaivable right. I do not mean, of course, that one cannot consent to (limited) silence or to being shielded from information. I mean that one cannot forfeit at least some aspects of the right to free speech conceived as the right to free communication and the right to free thought. While I think that an outright forfeiture of the active right to free speech could have devastating effects for the right-holder's well-being, it seems to me that the case for the unwaivability of the passive right to free speech is much stronger in light of Shiffrin's analysis. Not only is it practically impossible to waive our right to think for ourselves without turning into something other than a human being, it is also morally indefensible. Consider the rationale for the unwaivability of the right not to be enslaved: the freedom to do what one wishes with one's body stops making sense once freedom itself is lost. In terms of personal well-being, the deciding factor here is not that a slave leads, almost by definition, a bad life, but rather that slavery makes self-authorship impossible. Even if my decision-making is terrible and I would be better off with someone else making decisions for me, the fact that this is now a *way of life* for me means that this is not *my* life anymore. Even if I am not in shackles, I have forfeited the ability to choose my pursuits and my actions, as well as the responsibility that comes with these choices.

Forfeiting the right to think autonomously is akin to intellectual slavery: my thoughts about morality, friendship, politics, religion etc may come out of my mouth but are not truly mine. And in order to have thoughts that are truly mine, I need free access to information, as secured by the passive right to free speech. This is an aspect of the right that only an interest-based approach can capture in full. To acknowledge that there is an aspect of the right to free speech that is unwaivable is not, however, to say that the right must be exercised at all times or as much as possible. It is in our interest (that is, beneficial to our well-being) to maintain the right to receive information without undue interference, but we have other interests that may be harmed by listening to others. Speech that may cause direct harm (for example, speech that incites violence against us) is an obvious case but not necessarily the only one. Some types of speech (e.g. hate speech) can cause psychological harm, which may be in our interest to avoid. In this case, the question that presents itself is who and how should balance the two interests to determine which should take precedence. I will return to this point later.

[42]Raz (n 9) 391.

The right to online free speech

Everything I have said so far only aspires to show that it makes sense to understand the right to free speech as (i) consisting of the right to free communication and the right to free thought; (ii) protecting our interest in having such thought and communication as crucial (especially the former) elements of self-authorship and personal well-being, at least in autonomy-supporting cultures. This does not necessarily mean that there is no aspect of the right to free speech that can be grounded on something other than our interests. More generally, there is no reason to believe that the goodness of a right (or anything else) cannot stem from two different types of normative considerations. In crude terms, it may be the case that some exercises of free speech will not be in our interest (in the sense that they will not enhance our well-being) and yet remain capable of claiming some kind of protection. We may concede this without injury to the main point defended so far, namely that the core of the right to free speech is best understood in the way described above. We may also concede that the value of free speech is not adequately described by the language of rights. This is a critique that affects all moral rights when defended from a consequentialist point of view.[43] Even, however, if the language of rights doesn't capture everything that we value about them, legal rights are still a valuable tool not only for the description of complex concepts and relationships, but also for change in those relationships.[44] A legal right to free speech allows us to identify right-holders, duty-bearers, and possible limitations. A legal right to free speech that acknowledges freedom of thought as one of its components further enables us to separate the absolute, unwaivable aspect of the right from the aspect that can be justifiably waived. In the remaining of this section, I will briefly discuss how these thoughts affect the right to free speech online and why the approach favoured so far is particularly adept to explain and defend the exercise of free speech on the internet.

The European Court of Human Rights ('the Court') has made it clear that the provisions of Article 10 of the European Convention on Human Rights (ECHR) apply to the online sphere.[45] The crucial part of Article 10 for current purposes is the following (§ 1):

> 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.

While the ECHR includes the right of thought in a different article (Article 9), Article 10 encapsulates much of what we have discussed so far about the

[43]Richard Arneson, 'Against Rights' (2001) 11 Social, Political, and Legal Philosophy 172.
[44]Elizabeth Schneider, 'The Dialectic of Rights and Politics: Perspectives from the Women's movement' (1986) 61 New York University Law Review. 589.
[45]*Ashby Donald and Others v France* App no 36769/08 (ECtHR 10/01/2013).

right to free speech: it includes both the active and the passive aspects of the right and it makes reference to opinions, which are, of course, the product of thought. In fact, the protection of opinion online is broad enough to cover 'likes',[46] an expression which lies somewhere between speech and thought. This level of protection is justified by the important role the internet -especially user-generated online content- plays in the dissemination of information.[47] Having said that, the Court's reasoning often departs from the interpretation of the right to free speech defended here, as it neglects the significance of freedom of thought besides Article 9. However, it seems to appreciate the significance of free speech beyond its value for democracy and particularly for self-development.[48] It is beyond the scope of this article to comment on the jurisprudence of the Court. The important point for current purposes is that, according to its caselaw, the legal right to free speech is protected online as it is protected in the physical world. My aim here is to discuss what this protection *ought* to be, but it is important to keep in mind that the legal duty of the state to protect free speech online is indisputable.

The caselaw of the Court notwithstanding, is the online world different in any way that could render everything I have said so far about the right to free speech indefensible or impractical? We may think that the legal right to free speech exists unchanged on the internet, but the latter offers unique opportunities both for expression and for its restriction. One obvious difference is the ability to speak uninterrupted: when one stands up in a park to make a speech, physical interference by others is likely; on the internet, one can be blocked or muted, but not entirely silenced by the listener. Is an interest-based approach somehow unsuitable for the online world? Quite the opposite, I think. If we want to protect online freedom of expression effectively rather than nominally, the interest-based approach is our only chance. The reason is, once again, the impossibility of unwaivability in the context of will theories. If our consent is enough to forfeit the right to free speech, we would severely undermine the power to claim its protection every time we click on 'I agree' or 'I consent' before using a website. However, the purpose of legislation such as the Online Safety Act can only be to go beyond this and protect the interests of internet users after they have consented to general rules and regulations. The rationale of this protection is that, for example, while adults can consent to the viewing of pornographic material, they can still be harmed by material that falls under the general description of pornography but is nonetheless not in their interest to view. Apart from a reminder of the questionable validity of many forms of

[46] *Melike v Turkey* App no 35786/19 (ECtHR 15/06/2021).
[47] *Delfi AS v Estonia* (2016) 62 EHRR 6.
[48] *Handyside v the United Kingdom* (1979-80) 1 EHRR 737 § 49.

consent, these concerns stem directly from the nature of online communication. Unless we care for the interests and well-being of the users, we have no reason to care about their right to free speech after they have consented to using a website. But we have very good reasons to believe that this is not always enough.

The online safety act: free speech, user interests, and 'legal but harmful' content

Two types of regulation of online expression are relatively uncontroversial: the protection of children from content that is suitable for adults and the protection of adults from directly harmful content. I say 'relatively uncontroversial' because one can challenge absolute restrictions for both groups. Children, for example, still have a strong interest in making some choices regarding the (harmless) information they receive (primarily to avoid indoctrination and secure the ability for future self-authorship). Adults, on the other hand, may be better served by risking some exposure to harm in order to further other interests (dangerous sports are a good example). Nonetheless, a general defence of justifiable restrictions on the above grounds can result, in mostly uncontroversial terms, in legislation that would render some instances of free speech illegal. The same cannot be said for content which appears to require regulation but does not fall in either category, and is, therefore, legal. In its earlier form, the Online Safety Act defined legal but harmful content (for adults) as content that would be viewed as offensive by a 'reasonable person of ordinary sensibilities'. This definition brings to the fore the uniqueness of the proposed provision. The ordinary sensibilities of reasonable persons would be the yardstick to determine what counts as offensive and, therefore, harmful speech that should be regulated. The focus on offence and the reference to ordinary sensibilities and reasonable persons raise familiar red flags. While harm is, more often than not, a measurable effect on personal well-being that most of us would prefer to avoid, offence is not. And, of course, even if there is a way to know what offends reasonable persons with ordinary sensibilities, it may still be in the interest of some of us to be offended in pursuit of a valuable goal. If it is the interests of users that we care about, this is something we should not overlook.

The Online Safety Act seems to acknowledge this need, as it makes explicit reference to the interests of internet users. Specifically, it refers to the protection of the interests of vulnerable adult users (S 65 (3) (c)); children (S 78 (2) (c)); those with protected characteristics (S 78 (2) (f)); individuals (S 149 (5) (b)); users (S 152 (3) (a)). Strictly speaking, the use of this language does not amount to an outright adoption of the approach I have been defending here. However, it is indicative of the need to include user interests in any

comprehensive description of the purposes of legislation such as the Online Safety Act. The guidance provided by governments or regulatory agencies like Ofcom regarding the content and clarity of the terms and regulations governing access to websites is useful but not the end of any discussion about user protection online. As already mentioned, an adult's consent to be exposed to pornographic material cannot possibly shield her from all potentially (legal but) harmful content. She may still be exposed to something that will cause her distress, trauma, or some other form of psychological harm. In order to truly protect internet users from harmful material, the legislator must look beyond consent to establish rules and practices that create duties of service providers to protect the interests of users. This duty of care can start with clear and accessible rules about what kind of legal content is allowed by each service provider,[49] but it cannot end there: as the government acknowledged, it is unreasonable to expect online providers to produce codes of practice for each category of potentially harmful content.[50] Consequently, the companies that provide online services would have considerable discretion in deciding what constitutes legal but harmful content, especially when the potential victims are adults.

The examples of legal but harmful content provided in the context of the discussion of the Online Safety Act were diverse. One category seemed to refer directly to the risk of physical or psychological harm caused by online bullying, intimidation in public life, or self-harm and suicide imagery.[51] Another category that is of particular interest included misinformation and disinformation. One example of such content is antivaccination advocacy against established medical advice[52] which does not intend to 'capture genuine debate'.[53] At first glance, these two categories present different challenges. Bullying and intimidation can cause psychological harm, easily identifiable by its symptoms. The crucial policy question here is not whether online users should have access to some kind of protection against such conduct but what constitutes bullying and intimidation. Admittedly, there is a grey area between expression that is mostly offensive and expression that is abusive. Self-harm and suicide imagery can also cause psychological harm, and this provides good grounds for regulation. It is worth noting that, interestingly, the impact of the internet on the mental

[49]Online Harms White Paper – Initial consultation response, Ch 2. Available at: https://www.gov.uk/government/consultations/online-harms-white-paper/public-feedback/online-harms-white-paper-initial-consultation-response
[50]Ibid.
[51]Ibid.
[52]Online Harms White Paper: Full government response to the consultation, Part 2 https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/online-harms-white-paper-full-government-response
[53]Statement by the Secretary for Digital, Culture, Media and Sport made on 7 July 2022 https://questions-statements.parliament.uk/written-statements/detail/2022-07-07/hcws194.

well-being of users is not as devastating as such legislative initiatives seem to suggest. On the contrary, the widespread adoption of internet use and online platforms in the last two decades is not consistently linked to negative outcomes.[54] If these empirical findings are accurate, the case for regulation loses traction, but, in principle, we may still have reasons to consider it. Finally, misinformation and disinformation pose an entirely different question: while they may cause considerable public harm (for example, by convincing many people to avoid vaccination), they do not do so directly. Keeping in mind that my focus here is on adults, what comes between receiving this information and not vaccinating is the *decision* not to vaccinate. This is normatively crucial, especially if the right to free speech protects us *qua* thinkers.

As all parties and stakeholders have repeatedly pointed out, the protection of online free speech is of paramount importance and one of the objectives of a defensible legislative framework. We are, therefore, faced with a conflict of intuitions and values: on the one hand, we want to preserve the right to free speech; on the other hand, we acknowledge that some types of free speech can cause harm. When this harm is clear and direct, the Harm Principle is activated, and the restriction of free speech is prima facie justified. But when the expression in question does not fulfil the requirements that would make it illegal, the protection of listeners is difficult to justify. It is, then, no surprise that the provisions regarding legal but harmful speech were removed from the Online Safety Act. However, as mentioned in the introduction, my goal here is not to focus exclusively on the Act, but to use it as a springboard for further thought on the very concept of legal but harmful content. In fact, much of the debate around this provision focused on the active right to free speech: regulating legal speech could be a slippery slope leading to the suppression of free debate and the censorship of the press, given that both exceedingly take place online. Despite legitimate concerns about equal standing in the public discourse,[55] it is true that exemptions for journalistic or 'democratically important' content (both included in the early drafts of the Act) can reduce the harm to pluralism and democratic debate. Furthermore, one of the most defining characteristics of the internet is the availability of a wide range of options. During the COVID-19 pandemic, many social media platforms regulated their content to protect the public from disinformation, but this only resulted in the migration of the affected users to other platforms. It seems, then, that the most important harm caused by such a legislative framework is not to be found in the restriction of the active right to free speech: unless we are

[54]Matti Vuorre and Andrew K Przybylski 'Global Well-Being and Mental Health in the Internet Age' (2023) Clinical Psychological Science.

[55]Kyle Taylor and Elen Judson, 'Exemptions, exceptions and exclusions: Why the Online Safety Bill protects disinformation and abuse over freedom of speech and journalism', (2022) Demos https://demos.co.uk/wp-content/uploads/2022/05/Exemptions_Exceptions_and_Exclusions___OSB___vF.pdf

willing to argue that a speaker ought to have access to *any* platform, having access to *some* platforms may be enough to protect her right to express her opinions, communicate, and seek validation for her views (i.e. her active right to free speech). It may not, if no such platform is available, but this will not be the rule on the internet. It is much more pressing to look at the effects of the regulation of legal but harmful content on the passive right to free speech. In order to be thinkers and decision-makers, we need to have access to information – it is in our interest. Some information may be harmful or potentially harmful to us and, of course, it is also in our interest to avoid harm. The conception of free speech that focuses on human beings as thinkers has, as I argued above, an unwaivable aspect, but it is perfectly consistent with refusing to receive information or speech that may be harmful. However, once this decision is made by someone else for our own good, the arrangement becomes vulnerable to a charge of paternalism.

Paternalism

Paternalism is a complex concept. Most people can recognise the most prevalent paternalistic laws or policies (such as seatbelt laws) but find it difficult to offer a definition. The most practical approach in the search for an uncontroversial working definition is to start with Mill's antipaternalistic statement in his famous Harm Principle:

> [t]he sole end for which mankind are warranted, individually or collectively in interfering with the liberty of action of any of their number, is self-protection. That the only purpose for which power can be rightfully exercised over any member of a civilised community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant. He cannot rightfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinion of others, to do so would be wise, or even right[56]

The objective of the 'simple principle' as Mill called it, is revealed in the last period: individuals ought to be protected from interferences that aim to restrict their freedom for their own good. It does not matter what kind of 'good' the paternalist has in mind (whether the goal is to make one happier or a better person). When broken down, Mill's principle seems to consist of three elements: a restriction of liberty of action; for the good of the person whose liberty of action is restricted; without that person's consent. If we replace 'liberty of action' with the term 'autonomy' to account for subtler paternalistic interferences that Mill could not have considered (for example, internet firewalls that render certain websites

[56]Mill (n 7) 8.

inaccessible without the user's knowledge that there was an 'action' for her to undertake in the first place), a hopefully uncontroversial definition of paternalism would take the following form: 'paternalism is the restriction of an individual's personal autonomy for the good of that individual'.[57] I use the term 'personal autonomy' to reiterate the distinction between autonomy as a conception of self-authorship (and, in a way, freedom) and moral autonomy in the sense of moral self-legislation.[58] In this essay, I use the term autonomy in the former sense only.

The comforting simplicity of this definition does not prepare the reader for the complexity of paternalism in all its manifestations. From a practical viewpoint, it is not always easy to detect the motives of legislators and policymakers. Take smoking bans as an example: while there is a clear risk to others (because smoking indoors creates a health risk for non-smokers that was not there before), there is little doubt that the policy is good for smokers as well, since they inevitably smoke less. We may also consider the prohibition of certain dangerous substances such as poisons: the law aims to protect consumers who might recklessly buy these substances, but its coercive force is mostly directed towards the seller, who will suffer the most severe consequences of an illegal sale. To address these complexities, theorists have introduced distinctions (such as pure/impure, direct/indirect, soft/hard paternalism), some of which will be useful in the course of this discussion. From a normative viewpoint, the most puzzling question is that of paternalistic means: what types of restriction count as an 'interference' and a 'restriction of autonomy'? Coercion is an obvious answer, especially for legal paternalism, since most paternalistic laws introduce coercive threats: 'if you do not fasten your seatbelt, you will receive a fine'. Manipulation also restricts autonomy, as it distorts the way we understand our choices to the extent that our decision may not be ours in any sense that is meaningful for self-authorship.[59] Compulsion is uncommon in practice, as it is not a legal tool, but we would not doubt that physically preventing an adult from swimming in rough waters constitutes paternalism. In most cases, there will be a law backing up the compulsion used by the relevant authorities (in this example, the actions of a lifeguard pulling a swimmer away from the sea). However, compulsion and quasi-compulsion are still available to the paternalist: building a fence to prevent access to a dangerous site or a firewall to prevent access to a website are familiar examples. Finally, the paternalist may simply make an option unavailable: a government that considers professional boxing dangerous may refuse to recognise boxing unions and approve the organisation of boxing events, so that professional boxing

[57]Kalliris (n 13) 25–26.
[58]Immanuel Kant, *Groundwork of the Metaphysics of Morals*, Mary Gregor (ed) (Cambridge University Press 1997).
[59]Raz (n 9) 377.

will simply not exist. Let us call this paternalistic strategy the (non-coercive and non-manipulative) removal of options and this type of paternalism *decisional* paternalism (because someone makes a decision for the individual, for her own good).[60]

The most defining element of paternalism is also the most important one for current purposes. In short, paternalism embodies a conflict of values and intuitions: on the one hand, there is freedom; on the other hand, there is personal well-being.[61] The obvious value of personal well-being makes paternalism worthy of consideration in the first place: the paternalist restricts our autonomy but claims to be benevolent, to care about our interests. Given the strong connection between rights and freedom, paternalism is directly relevant to the restriction of certain rights, when that restriction aims to benefit the right-holder. One approach, commonly associated with liberalism, favours freedom to the extent that all paternalism is considered presumptively blameable.[62] This view sits well with most will theories of rights, as the choice (or lack of consent) of the right holder should be enough to render any paternalistic interference unjustifiable. For an interest-based approach, like the one defended here, it is usually a question of balancing the values in question: as important as personal well-being is, autonomy must be protected as well. This balancing can be a complex enterprise for those who believe, as I have argued here, that autonomy is indispensable to well-being. Elsewhere, I defend a balancing test which I think encapsules much of what many people, even those who believe that autonomy is intrinsically valuable, regard as justifiable paternalism. I conclude that 'paternalism is only justified when it seeks, through minor and mild restrictions of autonomy, to protect a threshold level of those basic goods that are components of the good life and, at the same time, secure the conditions for both autonomy and the successful pursuit of projects'. If these requirements are not met, autonomy should outweigh any concerns for personal well-being (this is what I mean by the term 'good life').[63] This test offers some clarity by reminding us of two -hopefully- uncontroversial facts about the two conflicting values: first, that not all goods (and, similarly, rights) are equally important for our well-being; second, that paternalism can harm the very thing it professes to protect because restrictions of autonomy often have an impact on self-authorship itself, and, ultimately, well-being. I submit that a similar balancing will be required in the context of the protection of the interests of internet users against harmful but legal speech. What is crucial for this discussion is to keep in mind that: (a) any interest-based approach

[60]Kalliris (n 13) 38.
[61]Kalliris (n 13) 24–25.
[62]Joel Feinberg, *The Moral Limits of the Criminal Law Volume 3: Harm to Self* (Oxford University Press 1989) 4.
[63]Kalliris (n 13) 44.

will require this kind of balancing; (b) any restriction of autonomy on paternalistic grounds requires special justification, due to paternalism's adverse effects on both autonomy and well-being.

The need to reconcile our love for freedom and our desire to protect personal well-being is so strong that one such attempt gained remarkable traction in recent years. In an influential book,[64] and several articles that followed, Richard Thaler and Cass Sunstein presented *libertarian paternalism*. Their theory was revised slightly in the years that followed, but the basis remained the same: libertarian paternalism uses insights from the behavioural sciences to create non-coercive interventions (called *nudges*) that can lead individuals to good decisions without restricting their autonomy. To do so, the nudger relies on *choice architecture*, i.e. the arrangement of people's choices so that they end up making the right choice (again, the right choice *for them*). The quintessential example of choice architecture and nudging comes very early in the book: a cafeteria manager who knows that people tend to select what they see, can arrange foods in a way that nudges people towards healthy choices, by placing unhealthy foods on the top shelves.[65] The customer is not coerced into eating healthily, and therefore, we hit two birds with one stone: we protect her well-being without restricting her autonomy – after all, the unhealthy choice is still there. Nudges have attracted criticism for being manipulative[66] or a path towards more paternalism, since there is no learning involved in choice architecture.[67] I will return to this point later. What is important to note is that nudging can be a powerful tool for choice architects whose objective is to direct internet users towards specific choices. In fact, the internet is an excellent platform for choice architecture, often with a focus on user safety.[68]

Online freedom of speech, user interests and the 'legal but harmful' debate

The regulation of legal but harmful speech on the internet could lead to a potential restriction of both the active and the passive right to free speech. While the right to express ourselves online does not directly require access to any specific platform, it does presuppose at least some prospect of communication. Many types of speech, especially those that express opinions

[64]Richard E Thaler and Cass R Sunstein, *Nudge: Improving Decisions about Health, Wealth, and Happiness* (Yale University Press 2008).
[65]Ibid 1.
[66]Till Grüne-Yanoff, 'Old Wine in New Casks: Libertarian Paternalism Still Violates Liberal Principles' (2012) 38(4) Social Choice and Welfare 635.
[67]Ricardo Rebonato, 'A Critical Assessment of Libertarian Paternalism' (2014) 37 Journal of Consumer Policy 357.
[68]Alessandro Acquisti et al., 'Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online' (2017) 50(3) ACM Computing Surveys 1.

about political, moral or existential issues, are part of our self-authorship and, ultimately, our well-being. A staunch proponent of human rights, for example, will find it difficult to pursue her goals without the *possibility* of communicating her views to others, both to achieve specific goals and to seek validation for this important life choice. The 'communication' aspect of the right to free speech can be devastated if the speaker practically has no ability to communicate her most life-authoring views. Recent developments reinforce the assumption that the internet will offer adequate opportunities for legal speech to become public. The COVID-19 pandemic and the 2020 US presidential election forced some social media platforms to ban users who repeatedly posted misinformation or disinformation, but most of them found alternative forums to express themselves. It can be said, therefore, that the vast majority of speakers had access to adequate online options, provided that their speech was not illegal. One objection against this argument is that these arrangements tend to create 'echo chambers', where the quality of the debate is significantly lower. This is a valid point, but it seems absurd to say that the active right to free speech includes access to all platforms or to the best forums for each debate. It is, of course, conceivable that some speakers will, in some circumstances, have no access to any platform at all. In this scenario, they would have a legitimate claim, provided that their speech is legal.

The preceding remarks do not entail that online service providers are justified in banning users on the grounds that they can always migrate to another website. A genuine interest-based understanding of the right to free speech acknowledges the value of the resources individuals dedicate to their pursuits, especially those which are central to their self-authorship. Internet users have, therefore, a valid claim to access to clear, non-arbitrary and reliable guidelines that ensure a level field of play, especially on those online platforms that tend to host important debates. This clarity is important, because free speech is part of self-authorship and well-being: we should be able to determine how to use our time and resources towards our goals. If my goal is to campaign for a presidential candidate, I must know the rules of the political debate on a social media platform before I begin to build my audience and establish myself as a political commentator. The same kind of clarity, however, is required on the part of the legislator. One of the most problematic aspects of the 'legal but harmful' approach is its conflict with the Harm Principle, which, as we saw, proclaims that wrongful harm to others can, other things being equal, justify a restriction of autonomy, but all other types of conduct, including non-wrongful harms and offence, must not. Some instances of online speech (hate speech, bullying, harassment, incitement to violence) can cause physical or psychological harm and can be classified under 'wrongful harm'. Misinformation and disinformation, on the other hand, can cause indirect harm, by encouraging internet users to make poor decisions. While this can be seen as wrongful behaviour,

its restriction would require the legislator to endorse the view that competent adults are not capable of decision-making. This is very clear in the example of vaccination: someone who chooses not to vaccinate makes this decision despite having access to plenty of advice to vaccinate. Therefore, a ban on anti-vaccination propaganda would describe the speaker as something that she is not -i.e. the main cause of a potentially harmful decision. Without clear principles guiding these assessments, the legislator can undermine the active right to free speech by refusing to draw a clear line between the kind of potentially harmful behaviour that justifies a restriction of autonomy and the kind that does not.

The passive right to free speech is more directly affected by the regulation of legal but harmful expression on the internet. Remember that this aspect of the right requires assuming both the role of the listener (in the context of a 'communication') and that of the thinker. I argued earlier in this essay that the latter is normatively and practically unwaivable, especially if we take the broader view which sees individuals not merely as contributors to democratic deliberation, but as moral agents who must think and act in many important areas of life. It follows that, while we do not forfeit the right to passive free speech by refusing to be a listener in specific contexts (e.g. because one detests racist views), the right can be severely undermined in two ways: when we refuse to be a listener *tout court*, in the rare circumstances where this will be possible; and when others decide what we should listen to (this is what I described earlier as decisional paternalism). In practice, it is common for others to decide what information will come our way: parents do it very often to protect or indoctrinate their children. For adults, however, this type of indirect paternalism is morally problematic. A paternalistic intervention is indirect when the paternalist relies on a third party to restrict the autonomy of the paternalized. Indirect paternalism seems prima facie less objectionable: Mill thought that a ban on private gambling would be unjustifiable, but the closure of gambling houses would not be.[69] However, there are two pressing objections to indirect paternalism, when directed at adults. First, it fails to respect them as autonomous agents by not addressing them directly: after all, a coercive threat, much like the ones commonly issued by the criminal law, assumes that the citizens are rational enough to appreciate the risk of breaking the law. This risk assessment may even result in accepting punishment to serve a goal that is crucial for our self-authorship. Second, entrusting a third party with the implementation of a paternalistic rule removes this special relationship of benevolence between the state and the citizen that can make paternalism more palpable. Unless the legislator can provide very specific guidelines to the third party, they jointly undertake the role of the paternalist (as part of

[69]Mill (n 7) 83.

their 'referee' role mentioned in the introduction). And while we may be willing to trust the legislator as someone who cares for our well-being, the same cannot be said about third parties, who have their own motives and interests. Something as simple as 'do not sell this chemical' does not transfer any real discretion to the third party, but general guidelines about what content can be classified as 'legal but harmful' does.

Since it is not possible to provide online service providers with specific and binding guidelines about the regulation of legal but harmful speech, the legislator shares with them the role of the paternalist, in an arrangement that can be described as indirect decisional paternalism. Even if we concede that governments can delegate decisions with direct impact on personal well-being to individuals or corporations, there are no guarantees that these providers will be motivated by the benevolence that is inherent in paternalism. Even those who take their duty of care very seriously are bound to look at it through the lens of their own pursuits. This could certainly be harmful for democratic debate (and another indication of the special role of free speech for democracy), but this is not the only area of concern. A thinker must form opinions about many important issues and the internet is a readily accessible source of information for a number of significant decisions. Vaccination is, again, a good example: whether we should be vaccinated or not, especially during a pandemic, is a decision with moral (it affects others) and political aspects. This is a decision that the individual ought to make for herself (including choosing to suffer justified legal sanctions, for example). The manipulation of the information to which she has access, even if it is for her own good, usurps this process. One possible response is that listeners can migrate to other platforms just as easily as speakers. But this is not true: while a speaker will actively seek an appropriate platform, a listener must grapple with so much information that the addition of new platforms would add even more to the noise through which she must navigate. Furthermore, the creation of echo chambers through migration may not harm the interests of the speaker to a significant extent, but, as already mentioned, the same cannot be said for the listener. A thinker must assess information and evaluate arguments. Echo chambers are not the appropriate place for debate. The ban of former President Donald Trump from Twitter (currently X) did not stop him from expressing his views, but it removed one source of information for the users who had to think about their vote. There is no doubt that it is sometimes in the interest of the listener to not have access to speech that may cause her distress, pain or simply confusion. But this is an assessment that she must make, aided by tools that allow her to stay away from that speech.

This brings us to the means available to the paternalist. There is always coercion – that is, a coercive threat that attaches a sanction to a specific act. This is a common tool for offences committed online – but not for legal content. As expected, the Online Safety Act offered, before its revision,

more than one options to online service providers. Apart from stronger responses, such as bans and 'take downs', providers could also rely on other forms of moderation, including effective warnings and providing users with tools to self-curate. Bans and take downs constitute what I described as a 'removal of options'. Removing an option for the good of the chooser may be more autonomy-restricting than manipulation: when we lie to someone (this would be a paradigm case of manipulation), we distort her perception of the world, but at least she may still be able to detect the lie and reinstate her initial, undistorted, set of options. The removal of an option leaves little room for the reinstatement of autonomous choice, as, more often than not, the chooser will not be aware that the options ever existed.[70] Banning a user or taking down a post amounts to removing an option without allowing the listener to assess whether it is in her interest to receive the removed information. In line with what I have already discussed above, if the content is illegal, there is always a good reason for its removal. However, if the content is legal, the listener ought to be allowed to exercise her passive right to free speech as an autonomous thinker.

An alternative to directly autonomy-restricting interventions such as bans and take downs would be, as initially intended in the Online Safety Act, to empower users in ways that could allow them to avoid undesired (harmful) content. This milder approach would require, at least in most cases, some form of choice architecture: warnings, hidden posts the user can choose to see, and community notes which rearrange the user's choices. They do so non-coercively and by maintaining the option to access the content in question and, therefore, are best understood as instances of libertarian paternalism: all these measures can operate as nudges. This type of nudging often creates defaults: once a certain post or user has been flagged as harmful, the listener is in fact asked to opt out from an established protective status. Defaults are problematic because they do not address us as autonomous thinkers but rather count on inertia to effectively commit us to a preselected option – and this inertia is very strong.[71] Furthermore, defaults specifically assume that we are not always thinkers. In Kahneman's famous terminology,[72] sometimes we make choices using System 1, which is fast, automatic and intuitive; sometimes we use System 2, which is slow, calculative and deliberative. Defaults are nudges that target System 1[73] and this directly undermines the passive right to free speech from a thinker's perspective. In short,

[70]Kalliris (n 13) 37. This does not suggest that a larger number of options is always better for autonomy. Certain arrangements of our options may make it easier for us to choose what we truly want (see Arvanitis, Alexios, Konstantinos Kalliris, and Konstantinos Kaminiotis, 'Are defaults supportive of autonomy? An examination of nudges under the lens of Self-Determination Theory' (2022) 59(3) The Social Science Journal, 394–404).

[71]Eric Johnson and Daniel Goldstein, 'Do Defaults Save Lives?' (2003) 302 Science, 1338.

[72]Daniel Kahneman *Thinking, Fast and Slow* (Penguin 2012).

[73]Cass R Sunstein, 'People Prefer System 2 Nudges (Kind Of)' (2016) 66 Duke Law Journal, 121.

certain warnings or notes tend to produce specific intuitive reactions, many of them related to biases like inertia or authority bias. The choice architect (in this case the service provider) is aware of these mechanisms and, therefore, can create powerful defaults from which most users will never opt out. This effect can be produced by entirely benevolent choice architecture, including community-based moderation, but is more powerful when backed by an authority like the government (or the law).[74] If we value free speech in the sense defended here, we must acknowledge that even these mild interventions can be autonomy-restricting and, ultimately, an unjustifiable paternalistic restriction of free speech. As in most cases of paternalism, the balancing between autonomy and well-being is left to the paternalist, thus undermining the ability of users to decide whether it is in their interest to risk exposure to potentially harmful content that is, nonetheless, legal and perhaps useful to them as thinkers. Consistently to the approach adopted here, one way to protect internet users without throttling their self-authorship would be to focus not on inevitably heterogenous groups but on individuals *qua* thinkers. This would mean, first, that thinking should be encouraged: any information shared in the form of warnings, community comments and other quasi-defaults must be accompanied by enough time (akin to cooling off periods) for the user to reflect on them. For example, users could be provided with substantive information about the post in question and be prompted again in the near future to make a decision. Second, when a more robust approach is deemed necessary (because, for example, the information in question is known to cause distress to most users), the way this is communicated and presented should contribute to learning. Even bans and take downs would, in this scenario, be potentially justifiable measures, provided that the (psychological) harm risked is severe and likely for the vast majority of users *and* the procedure is transparent enough to allow users to reflect on the merits of the decision and, if they so decide, seek the removed information elsewhere.

Conclusion

The regulation of legal but harmful online content aimed at addressing a real problem: in their pursuit of information which will allow them to make decisions as autonomous thinkers, users can come across content that may cause (mostly psychological) direct harm, distress that does not amount to harm or poor decisions through disinformation/misinformation. If we understand the right to free speech as allowing the right-holder both to communicate and to operate as an autonomous thinker, it becomes clear that the

[74]Craig McKenzie Craig et al, 'Recommendations Implicit in Policy Defaults' (2006) 17 Psychological Science, 414.

latter is an aspect of the right we cannot forfeit. To be thinkers in this sense, what I described as the passive right to free speech is essential: we must receive information without undue interference or manipulation from third parties. As we exercise our passive right to free speech, we risk exposure to harmful content. When this content is directly and wrongfully harmful, the Harm Principle takes over and regulation is justifiable. But when the content is not directly and wrongfully harmful (and, therefore, prima facie legal), the balancing of autonomy and well-being becomes much more complex because it is no longer about the autonomy of the wrongdoer versus the well-being of the victim. In this context, it is the same person whose autonomy and well-being are at stake, and, consequently, any intervention would be paternalistic. Bans and take downs remove sources of information, making decisions for the thinker without her involvement. They are the most problematic measures, especially when service providers are given discretion to exercise indirect paternalism. If, however, a case can be made that they are likely to cause psychological harm to a large number of users, their use could be justified provided that they allow affected users to reflect and learn. Choice architecture by means of warnings, hidden content and community notes can also undermine (passive) freedom of speech, since nudges of this kind tend to exploit biases rather than treat users as thinkers. And, as we saw, it is not as easy to migrate to other platforms for the listener as it is for the speaker. They are more easily justified when they leave room for consideration. Learning, and informed decision-making for internet users.

Acknowledgement

I am grateful to Hedvig Schmidt, Michael Da Silva, Napoleon Xanthoulis, the participants in the Centre for Law and Technology Research Seminar Series, and two anonymous reviewers for their valuable feedback.

Disclosure statement

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

Notes on contributor

Konstantinos Kalliris is a Senior Lecturer at Essex Law School. His research interests include legal and political theory, criminology, and the behavioural analysis of law. His published work focuses on various themes, including paternalism, personal autonomy, organised crime, and the study of judicial systems.

ORCID

Konstantinos Kalliris http://orcid.org/0000-0003-4566-1681