Hate Speech and Public Reason

Angelo Ryu*

Abstract—This article presents a novel defence of the free speech principle. It is best understood, not as an individual right to speak freely, but as parasitic to the requirement of liberal states to conform to the requirement of public reason. As such, it protects our interests, not *qua* speaker or listener, but as participants in constitutional government. This account challenges the orthodox view on the legal protection of hate speech. It suggests that verbal hate speech deserves special protection, but nonverbal speech does not. The article concludes by comparing the legal status of hate speech in the United States, the Council of Europe, and England and Wales. Each suffers from varying degrees of normative deficiency.

* St John's College, Oxford. Many thanks to the editorial team at the Oxford University Undergraduate Law Journal for their patience and helpful suggestions. All errors and omissions remain my own.

Introduction

To what idea does the free speech principle refer? At bottom, it is this: there are many reasons to regulate action. But there is something special about speech, and to regulate it requires special justification.

There is widespread agreement on two points. First, the free speech principle protects a valuable right. Second, the principle is subject to limits. There is less agreement, though, on the correct approach to take in determining those limits. At the margins, the scope of the principle is heavily contested. One contentious area is the legal status of hate speech. It targets the most vulnerable among us and causes serious harm. How far, then, should the law protect hate speech as free speech?

Let us begin with a working definition of hate speech. For our purposes, the term refers to speech which stigmatises on the basis of a protected characteristic, like race or religious affiliation. What unites this category of speech is its expression of contempt towards the status of a class of persons in society.

How, though, might one express such contempt? Broadly speaking, I suggest the available means of communication fall within two camps: verbal and nonverbal speech. The former relies on the use of language, while the latter does not.[1] By this I refer to the intuitive distinction between, for instance, the written words of an opinion column and the act of burning a newspaper. Another example might be the difference between the televised political debate and the wearing of black armbands to protest the Vietnam War.

[1] To be clear, speech is verbal if it is language-based, even if it is not written or spoken. For our purposes, there is no relevant difference between a hateful comment which is spoken as opposed to delivered in sign language.

What, though, is special about language? The relevant distinguishing feature, I suggest, lies in its capacity to supply reasons which purportedly justify the claim being made. That is, the use of language allows for, but does not necessarily entail, a public interchange of reasons for belief. Nonverbal speech, on the other hand, is mostly limited to signaling support for a certain position. As such, they differ in their capacity to facilitate a process of public justification.[2]

The relevance of this distinction is often overlooked. This is because expressive freedom is commonly justified on grounds which have nothing to do with the communication of reasons. For example, this distinction would have no purchase if the value of free speech lies solely in people having the opportunity to express their identity, or the opportunity to signal support for a certain position.

This article challenges the orthodox view. Indeed, I argue that the free speech principle, as a distinct idea, is justified precisely because it protects the public exchange of reasons. The principle reflects the limits placed by the requirement of public reason on the state's freedom of action.

As a practical matter, this account of the free speech principle has important implications for the regulation of hate speech. It draws the boundaries of acceptable regulation differently from those which exist in the positive law of various legal systems.

The argument proceeds as follows. First, I develop an account of free speech based on public reason. Second, I explain how this account leads to a principle requiring the protection of

[2] Importantly, the distinction relates to capacity, not capability: it is possible to have verbal communication which does not communicate reasons, and to have non-verbal communication which tacitly communicates reasons.

certain forms of hate speech. Third, I consider two objections to the principle. Finally, I take a comparative approach and evaluate the regulation of hate speech in the United States, the European Court of Human Rights, and England and Wales.

1. Pluralism and Public Reason

In pluralistic societies, there is often serious and entrenched disagreement over fundamental moral principles. This leads to a puzzle: how might a state justifiably enforce moral principles in the face of fundamental disagreement while remaining tolerant of diverse views?

Although universal consent resolves the problem of disagreement, it is too restrictive a requirement to allow for practical governance. A state could accomplish little, if anything at all, if it needed to secure the agreement of all its citizens before it could act. Alternatively, a state might rely on a perfectionist claim to moral truth. On this view, a state is justified in enforcing a moral principle if that principle is true. While this approach resolves the problem of practical governance, it fails to respect the presence of fundamental disagreement over the truth of certain moral claims.

For a while, the predominant move in political philosophy has been the adoption of a requirement for 'public reason', or the idea that the state must act in ways justified with reasons everyone can reasonably accept.[3] This is a middle ground between consent and truth. It is distinct from the former since it does not require that everyone actually agree to the specific policy. All that is required is that the policy be justified in a way reasonably acceptable to everyone. It is distinct from the latter

[3] John Rawls, *Political Liberalism* (Columbia 2005) 217.

since it relies on the acceptability of the means of justification, as opposed to the truth of the conclusion.

The requirement of public reason is secured by the provision of reasons which are of a type capable of acceptance in the face of fundamental disagreement over comprehensive worldviews. Importantly, this does not ask whether all reasonable people agree that the supplied reasons justify the restriction. Rather, it limits the category of acceptable reasons to those which are not premised on the acceptance of a reasonably contestable belief.

The crucial claim in this article is that the free speech principle follows from the adoption of public reason as a restriction on state activity. As I understand it, the free speech principle points to a restriction on the *type* of reason capable of justifying the regulation of expression. Specifically, legal restrictions on the communication of reasons cannot be justified by the content of those reasons.

To see why, consider the distinction between practical and epistemic reasons. The former consists of reasons to do, or not do, an act. The latter consists of reasons to believe, or not believe, a claim.

As directed towards the state, the requirement of public reason is best understood as a practical reason. It is a reason for the state to limit its actions to those capable of being justified in a certain way. Specifically, state action must be justified in a way which allows for the possibility of obtaining a certain epistemic fact: a belief in the acceptability of that mode of justification.

That the requirement of public reason restricts the state's freedom of action is uncontroversial. What is controversial, however, is understanding the restriction as conditioned on an epistemic, as opposed to a normative, fact. This choice belies an

important distinction in how scholars have conceptualised public reason.

Recall that public reason restricts state action to those which can *reasonably* be accepted by all. Here, 'reasonably' is clearly doing a lot of work. What it allows us to do is assume certain favourable conditions about the world. This is what is known as ideal theory.

There is deep disagreement over the best way to idealise the requirement of public reason. A normative approach would idealise the theory by limiting the applicability of public reason to those who accept the fundamental tenets of liberal society. This could, for instance, include accepting the Rawlsian claim that society consists of a reciprocal system of fair cooperation.[4]

A purely epistemic approach would idealise the theory by limiting the applicability of public reason to those who are rational and capable of logical reasoning. Importantly, it is inclusive of a wide array of prior evaluative commitments. On this view, public reason applies to those who hold flatly divergent views about matters as fundamental as justice.

The difficulty with both approaches is that they risk being self-defeating. The normative approach is difficult because it risks collapsing into perfectionism.[5] Public reason is said to be required because of the problem of fundamental disagreement within pluralistic societies. A prior determination of acceptable and unacceptable moral claims, however, risks replicating the problem which the approach sets out to solve. For example,

---

[4] *ibid* 49.

[5] Joseph Raz, 'Disagreement in Politics' (1998) 43 *American Journal of Jurisprudence* 29-30.

Rawls famously claimed that the trimester analysis adopted in *Roe v Wade*[6] was incapable of admitting to reasonable disagreement.[7]

The purely epistemic approach, on the other hand, is difficult because it risks collapsing into consent.[8] By failing to bracket the limits of reasonable disagreement, the concern is that public reason will be incapable of coming to a consensus about anything of value. In short, the worry is that this approach renders public reason impotent.

Generally, those who go the epistemic route ask whether the state action is justified on grounds which all citizens, given certain assumptions of rationality, would accept after a 'respectable amount' of sustained thought.[9] In response to the impotence objection, however, we might want to relax this constraint. A promising way to do so, I suggest, follows from a feature of public reason which distinguishes it from perfectionism, namely its attempt at bypassing, without resolving, the truth of moral claims.[10] Specifically, we might move away from asking whether citizens 'would' accept the grounds of justification and, instead, ask whether they 'could'. In other words, public reason requires the state to act in a way which allows for the epistemic possibility of universal acceptance.

Importantly, this analysis leads to a thin conception of reasons. This is because reasons, in the sense relevant to the idea of public reason, consist of claims which aim to supply a justification *to others*. Schoelandt refers to this as interpersonal, as

---

[6] 410 US 113 (1973).
[7] Rawls, *Political Liberalism* (n 3) 243, fn. 32.
[8] Jonathan Quong, 'What is the Point of Public Reason' (2014) 170 *Philosophical Studies* 545-553.
[9] Gerald Gaus, *The Order of Public Reason* (Cambridge 2011) 250.
[10] Rawls, *Political Liberalism* (n 3) 94.

opposed to impersonal, justification.[11] The former consists of reasoning about the reasons which might appeal to others, while the latter consists of reasoning directly about the reasons as they actually exist.[12] As such, 'the mere fact that there exists a valid justification for true premises for some proposition Q does not mean that we have just justified Q to Peter'.[13] Regardless of whether Q is true, interpersonal justification depends on the ability of Peter, as an epistemic matter, to accept Q.

For this epistemic possibility to obtain, the state must allow for the public interchange of reasons. That is, the state must give its citizens the chance to convince others of their views, and to be convinced in turn. When the state acts for reasons which are rejected by certain members of its polity, it must leave open the possibility for those members to contest those reasons. This explains why the freedom of speech deserves special protection in a pluralistic society.

2. Protecting Hateful Reasons

There is something deeply unsettling about grounding the free speech principle in an account of public reason. So far as hate speech communicates reasons, they consist of paradigmatically non-public reasons. Despite this, the free speech principle emerges from the analytic priority of the need to justify an exercise of coercive political power.[14] That is, even if public

[11] C. Van Schoelandt, 'Justification, Coercion, and the Place of Public Reason' (2015) 172 *Philosophical Studies* 1031, 1033.
[12] cf Joseph Raz, *Practical Reason and Norms* (Oxford 1999) 17: 'reasons are used to guide behaviour, and people are guided by what is the case, not what they believe to be the case'.
[13] Jonathan Quong, *Liberalism without Perfection* (Oxford 2010) 142.
[14] Charles Larmore, 'What is Political Philosophy?' (2013) 10 *Journal of Moral Philosophy* 293.

reason confers duties on all of us,[15] the state cannot point to an individual's breach to absolve itself of the fundamental requirement of public reason. Put another way, while speakers might have a duty to refrain from saying hate speech, the state lacks the legitimate authority to enforce that duty, so far as it violates the free speech principle.[16]

Unsettling as it might be, there are compelling advantages to bringing the public reason literature in conversation with free speech theory. First, it situates the relationship between free speech and legitimacy in an overarching analytic framework. Second, it offers a more robust account of the value of free speech, one which considers the interests of both speakers and listeners.

By contrast, most scholars focus exclusively on one aspect of free speech. Let us first consider the speaker-based views. Some ground the value of expressive freedom on democratic participation. It allows citizens to plausibly claim authorship over the laws which govern them, since they had a voice in forming the majority will.[17] Others ground expressive freedom in individual autonomy, based on the value of 'self-

[15] Rawls, *Political Liberalism* (n 3) 217-18.
[16] cf Jeffrey W Howard, 'Dangerous Speech' (2019) 47 *Philosophy & Public Affairs* 244-47. In this article, however, I advance a reason to constrain the state's power of enforcement which can plausibly be understood as a 'trump', as opposed to an interest to be weighed in a proportionality analysis.
[17] Ronald Dworkin, 'A New Map of Censorship' (2006) *Index on Censorship* 131; Robert Post, 'Participatory Democracy and Free Speech' (2011) 97 *Virginia Law Review* 482.

fulfillment'.[18] It allows us to realise the things which 'correctly constitute the core of what we value about ourselves'.[19]

Let us now turn to listener-based views. A prominent variant of this argument revolves around a metaphorical 'marketplace of ideas'.[20] On this view, speech promotes the discovery of truth by preventing 'a sort of intellectual pacification'[21]. It allows people to hear 'everything worth saying'.[22] Others argue that expressive freedom contributes to the individual self-realisation of listeners.[23] By protecting a 'free flow of information',[24] expressive freedom allows listeners to make informed decisions.

In this article, however, I rely on the concept of public reason to build an account which recognises the value of free speech to speakers *and* listeners. Both participate in a valuable activity which deserves protection. Specifically, the content-based proscription of pure speech interferes with the ability to supply and consider reasons in the public sphere, and therefore prevents the public interchange of reasons. This account consists of two claims.

---

[18] C. Edwin Baker, 'Scope of the First Amendment Freedom of Speech' (1978) 25 *UCLA Law Review* 991.
[19] Seana Shiffrin, 'A Thinker-Based Approach to Freedom of Speech' (2011) 27 *Constitutional Commentary* 285-87.
[20] This view was prominently advanced by Justice Oliver Wendell Holmes: see his dissent in *Abrams v United States*, 250 US 616, 624 (1919).
[21] J. S. Mill, *On Liberty* (Hackett 1978) 31.
[22] Alexander Meiklejohn, *Free Speech and Its Relation to Self-Government* (Harper 1948) 25.
[23] Martin Redish, 'The Value of Free Speech' (1982) 130 *Pennsylvania Law Review* 593.
[24] *ibid* 604.

> *Speakers*: The regulation of hate speech constitutes an impermissible restriction on the liberty of speakers to present an argument in the public sphere.
>
> *Listeners*: The regulation of hate speech infringes on the ability of listeners to consider and evaluate the views of their fellow citizens.

These claims are often understood as separate arguments for limiting the regulation of hate speech. However, they are better understood as protecting different aspects of the same joint activity, namely the ability of individuals to participate in a process of public justification. This is something speakers and listeners do together.

A. Giving Reasons

There are many reasons to proscribe expression. First, the expression might be constitutive of an act which has legally significant consequences.[25] This includes saying 'I do' in the context of a wedding or 'You are fired' in the context of employment. Second, the expression might lead to a situation with legally significant consequences. An expression might cause, for instance, an untenable work environment or a disruption to public order. Third, the expression might communicate a message thought to be normatively wrong. In other words, the proscription of expression could be justified on the basis of its content.

As I understand it, the free speech principle restricts the type of reasons which the state might offer to justify the legal regulation of expression. Specifically, it prevents the state from relying on a content-based rationale. Thus, while the first and second reasons are open to the state under the free speech

[25] This point is explored further in subsection 3B.

principle, the third is not. This restriction follows from a distinctive feature of law, namely its exclusionary character.

By this I refer to Joseph Raz's account of law as it relates to how we reason through moral issues. To him, laws offer, inter alia, exclusionary reasons for actions.[26] This requires some unpacking. A first-order reason for action is a reason to do or not do an act. For example, a belief in helping others provides a positive first-order reason to give to charity. Second-order reasons are different. They are reasons for, or against, acting on a reason. According to Raz, the law supplies negative second-order reasons, which he calls exclusionary reasons. This means that the law gives us a peremptory reason to disregard conflicting reasons for action.

The free speech principle protects the expression of reasons which purportedly justify a belief in a claim. This is because the state, by proscribing the articulation of these reasons on the basis of their content, adopts an impermissible justification for its action. To clarify, a content-based restriction is one which is putatively justified by the wrongness of the expressed meaning. Such restrictions, when imposed by law, require the speaker to disregard other reasons she might have to express the proscribed content.

The expression of justificatory reasons communicates a particular type of message. Specifically, it communicates a message which purportedly consists of epistemic reasons to believe a certain claim. When the state proscribes the communication of this type of message on the basis of its content, it adopts an exceptional justification. It is exceptional because it prevents the speaker from explaining why she rejects the rationale

[26] Raz, *Practical Reason* (n 12) 35-48.

of the law in the public sphere. In other words, it prevents the speaker from communicating reasons to reject the law.

A law which prevents same-sex marriage restricts the ability of some to say 'I do'. It does not, however, prevent speakers from providing reasons for why they should be allowed to marry. A law which prevents the formation of a hostile work environment restricts the ability to say certain things within the workplace. It does not, however, prevent speakers from providing reasons for why a workplace does not deserve protection.

By contrast, hate speech laws restrict the ability of speakers to make certain arguments. Such a restriction is often justified on the content-based ground that they communicate a message which is wrong. For instance, the message might make a claim regarding racial inferiority. Here, the speaker cannot advance an argument which challenges the rationale underlying the law, which in this instance is racial equality. This is because denial of racial equality is exactly the thing which renders the expression hateful, and thus legally proscribed. As such, the speaker does not have the ability to publicly defend her views, and to therefore participate in a public interchange of reasons.

B. Considering Reasons

Next, let us consider the role of the free speech principle in protecting the ability to consider reasons. Many arguments for the free speech principle are grounded on the positive net benefits of protecting expression.[27] T.M. Scanlon, by contrast, advances a

[27] Many philosophers take this position. Eric Barendt, *Freedom of Speech* (Oxford 2007) 15 makes an instrumental argument for personal development. Joseph Raz, 'Free Expression and Personal Identification' (1991) 11 *Oxford Journal of Legal Studies* 310 argues for expressive freedom on the grounds that it validates lifestyles. This argument is also made in the legal academy. Alan Chen and Justin

non-consequentialist argument.[28] He begins with the premise that a legitimate government is 'one whose authority citizens can recognise while still regarding themselves as equal, autonomous, rational agents'.[29]

This account of autonomy precludes a legitimate state from preventing subjects from independently deliberating on the merits of an act. From this comes Scanlon's Millian principle, which he argues is a natural extension to Mill's harm principle. It consists of two harms.

> *Judgement*: harms suffered by an agent as a result of having a false belief, due to their having listened to another agent's act of expression.
>
> *Persuasion*: harms that come as a consequence of an agent who acts based on the mistaken belief, caused by the expression in question, that such an act is worth performing.

According to Scanlon, neither *judgement* nor *persuasion* can justify legal restrictions on expression.[30] Restrictions on expression based on *judgement* prevent listeners from forming false beliefs. Scanlon argues this deprives listeners of the right to independent judgment because she cannot rely on her own reasoning, as opposed to the state. Restrictions on expression

---

Marceau, 'High Value Lies, Ugly Truths, and the First Amendment' (2015) 68 *Vanderbilt Law Review* 1437-38 finds that expression instrumental to discovering illegality should be protected. Martin Redish, 'The Content Distinction in First Amendment Analysis' (1981) 34 *Stanford Law Review* 119-21, 136 argues that expression should not be regulated because free speech is instrumentally good for democratic governance.

[28] T.M. Scanlon, 'A Theory of Freedom of Expression' (1972) 1 *Philosophy & Public Affairs* 205.

[29] *ibid* 214.

[30] *ibid* 213.

based on *persuasion* prevent the advocacy of certain types of activity. Scanlon argues this deprives the listener of the opportunity to independently judge whether that advocacy ought to be acted upon. Thus, the Millian principle suggests that the regulation of hate speech may infringe on the *listener's* autonomy.

The difficulties with grounding a free speech principle on a freestanding conception of autonomy are well known.[31] One major difficulty arises from the possibility that hate speech might cause certain harms which arise unconsciously, and therefore in the absence of an exercise of the victim's independent judgement.[32] If so, it would be difficult to argue that the proscription of hate speech, in this instance, infringes on the victim's autonomy interests.

The Millian principle, however, becomes more plausible when understood in light of a broader framework of public justification. Put in this light, it stands for the proposition that listeners must have the opportunity to exercise their independent judgement and come to a conclusion regarding the publicly expressed reasons of others. In this way, my account avoids having to rely on autonomy as an end in itself. Rather, the victim's exercise of autonomy is understood as part of a valuable joint activity. Importantly, since the value of this activity is based on its contribution to the *possibility* of acceptance, it is not defeated by a specific instance in which a victim fails to exercise her independent judgement.

---

[31] Susan J. Brison, 'The Autonomy Defense of Free Speech' (1998) 108 *Ethics* 324-36; T.M. Scanlon, 'Why Not Base Free Speech on Autonomy or Democracy?' 97 *Virginia Law Review* 546.

[32] A good discussion can be found in Charles R. Lawrence III, 'The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism' 39 *Stanford Law Review* 317-88.

3. Two Objections

In this section, I consider two potential objections to my account. Both look to undermine the case for protecting hate speech by analogy. Specifically, they attempt to demonstrate that hate speech is similar to a type of speech whose regulation is unobjectionable. The first objection argues that hate speech, like false advertising, reduces the deliberative capacity of its listeners and should be regulated on that basis. The second objection argues that hate speech, like saying 'I do' at a wedding, alters the normative status of others and should be regulated as a speech-act. These objections, however, fail.

A. False Advertising

First, let us consider the false advertising objection. The argument goes as follows. If the free speech principle is justified by the value of autonomy, one might think that autonomy is something which the state ought to encourage. But if autonomy is a good which can be increased, as opposed to a binary concept, it follows that issues of autonomy are 'a matter of degree'.[33] If so, it is not at all clear whether we should protect acts which are 'a product of social conditions that fail to allow for autonomy, rightly understood'.[34] Thus, if hate speech misleads others to act on wrong reasons, it ought to be regulated on the same basis as false advertising.

As a preliminary matter, I note that legal restrictions on false advertising and hate speech are distinguished by their scope. The former regulates expression for the *purpose* of selling goods or services.[35] Such advertising is regulated to protect a certain act,

[33] Joseph Raz, *The Morality of Freedom* (Oxford 1986) 156.
[34] Cass Sunstein, *Democracy and the Problem of Free Speech* (Free Press 1993) 138.
[35] Consumer Protection from Unfair Trading Regulations 2008, s2(1).

namely informed buying in the teeth of an unequal bargaining position.[36] It would not, however, prevent a seller from explaining or defending their advertisement through, for instance, a press release. By contrast, legal restrictions on hate speech regulate expression for the *purpose* of preventing the communication of its content. It therefore closes off all avenues of public justification.

More importantly, the regulation of false advertising is often justified by pointing to the deliberative capacity of listeners. In other words, these restrictions are justified because false advertising hinders the ability of consumers to make the right choices for them. For the false advertising objection to hold, it must demonstrate that the same reasoning applies to listeners of hate speech. This line of argument, however, does not translate well from false advertising to hate speech. From the standpoint of the consumer, discovering the falsity of advertising is difficult due to the unequal distribution of information. This argument, however, becomes weaker once we move from factual to evaluative statements. It is uncontroversial to point to an epistemic deficit in the case of false advertising. To regulate hate speech on this basis, however, assumes that the members of the polity are unreasonable, and therefore incapable of exercising reasonable judgement. This position denies the compatibility of public reason with democratic governance.

One might reasonably object and say that hate speech and misinformation often go hand in hand. That is, hate speech might be perpetrated through the distribution of false facts. It is therefore argued that restrictions on this type of speech are justified by listeners not having access to the relevant facts, as opposed to the unreasonable nature of their evaluative position.

Ultimately, though, this objection fails. A plausible justification for restricting false advertising is that consumers

[36] *ibid*, s3.

suffer from an epistemic deficit *by virtue of* their status as consumers, as opposed to sellers. For example, certain flaws about a product might only be revealed through sophisticated tests. The same does not hold for hate speech. Restrictions on hate speech are not justified because of the epistemic limitations of victims (such as consumers) as opposed to wrongdoers (such as sellers). Even if the hate speech is premised on falsifiable facts, their regulation as a discrete class of acts, and not false claims more generally, is based on the wrongness of the underlying evaluative claim. For instance, 'the Communists should be jailed because they burned the Reichstag building' might trigger special regulation because the implied link between political ideology and normative desert is wrong. This, though, is exactly the move precluded by a commitment to public reason.

There is an important limit to the scope of a free speech principle grounded in public justification. It lies in the distinction between viewpoint and viewpoint-neutral discrimination. The state, in directly targeting hate speech, engages in a form of viewpoint discrimination. It requires the state to take a side. In doing so, the regulation of speech reflects a prior judgement on the reasons which might be supplied to defend a viewpoint. This poses special difficulties in the context of speech restrictions. These problems, however, do not arise in viewpoint-neutral restrictions. Put simply, there is a difference between restricting speech because of its communicative content and restricting speech because it is necessary to prevent a separate harm.

We might formalise this distinction as follows:

> *Regulation of content*: The ability of a speaker to say certain things is restricted because of what the speech means.
>
> *Regulation of consequence*: The ability of a speaker to say certain things is restricted because of what the speech does.

The distinction turns on a difference in the way a speech regulation is justified. Importantly, this does not entail a strict conceptual division. For instance, the regulation of consequence might have a disproportionate effect on speech which expresses a specific viewpoint. This is because the harmful effect of speech may vary depending on the content of speech. Thus, speech proscriptions which have a discriminatory effect on viewpoint are not necessarily objectionable, so long as they are solely justified on viewpoint-neutral grounds.

As such, the state may regulate expressive acts *qua* acts. For example, a blanket ban on any expression which causes harm would solely target the act, and could therefore be justified on that basis.[37] So too can the regulation of defamation, which punishes the act of expressing a falsehood that causes reputational injury to another person, regardless of the specific viewpoint expressed, subject to various defences.[38]

This distinction helps explain why the regulation of hate speech is distinct from, say, the regulation of certain religious practices which involve the taking of illegal drugs.[39] Drug regulations protect against harm arising from the consumption of certain substances. This harm arises independent of the religious context. That is, drugs cause harm regardless of why the drug is consumed. The harm of hate speech, however, lies in the message it is intended to communicate.

---

[37] As it relates to the free speech principle. To be clear, I am not arguing that such a blanket ban would, all things considered, be justified.

[38] Defamation Act 2013, s2-7.

[39] Compare *Employment Division v Smith*, 494 US 872 (1990).

B. Speech-Acts

Next, let us consider the objection that hate speech constitutes a speech-act, and is therefore open to regulation independent of its expressive function. This line of argument is pursued by Maitra and McGowan.[40]

They argue that an 'obligation-enacting utterance' should not be covered by the free speech principle.[41] By obligation-enacting utterance they refer to expression which alters one's obligations simply by being said.[42] So, for example, the statement 'You're fired' alters obligations by terminating the employment relationship. The employer need not do anything further; saying the statement is enough. The expression constitutes the act of termination, which in turn is subject to legal regulation.

How, though, might hate speech alter obligations? For now, let us consider a provisional answer: by marking certain groups as inferior, hate speech imposes a social obligation to treat them as if they were inferior. For example, Charles Lawrence argues that hate speech is constitutive of, and has effect in, an unjust social hierarchy which systematically discriminates against racial minorities.[43] If so, a listener of hate speech might be placed in a rule-bound position. Maitra and McGowan acknowledge that this claim may be vulnerable to arguments challenging the speaker's authority to enact such a hierarchy.[44] It is one thing to

---

[40] Ishani Maitra and Mary McGowan, 'On Racist Hate Speech and the Scope of the Free Speech Principle' 23 *Canadian Journal of Law & Philosophy* 350.
[41] *ibid* 354, 369.
[42] *ibid* 350-54.
[43] Charles Lawrence 'If He Hollers, Let Him Go: Regulating Racist Speech on Campus' in M. Matsuda et al, *Words that Wound: Critical Race Theory, Assaultive Speech, and the First Amendment* (Westview Press 1993) 53.
[44] Maitra and McGowan (n 40) 369-70.

argue that hate speech perpetuates systematic discrimination, it is quite another to claim it *enacts* it.

However, even if we accept Lawrence's claim, this argument fails. When speech is solely a speech-act, it falls outside the free speech principle. While having the outward trappings of speech, it is not truly expression as it communicates no expressive content. Many times, however, speech both furthers an act and communicates expressive content. This, I imagine, is what Lawrence would claim about hate speech. It communicates a racist idea and enacts a discriminatory social hierarchy. Let us call this type of expression 'mixed speech'.

In most instances, the normativity of mixed speech is separable from the expressive content itself. It is, however, inseparable in the case of hate speech. Statements with a separable normativity have a normative force distinct from the issue at hand. An example of this is political obligation, where the obligation to obey the law is independent of the content of the specific law. In other words, the obligation acts, in some sense, as a 'trump' over our bare preferences.[45]

Compare this to inseparable contentions. The normative force of hate speech resides entirely in its expressive content. Hate speech relies on the normative claim that those of a certain protected characteristic deserve to have a lower social standing. It is this reliance on the normative claim which makes it hate speech. It is also the acceptance of that claim which obliges a bigot. It goes no further than this: whether the claim obliges is wholly dependent on how one feels about the protected characteristic.

Statements which depend on a normative claim separable from the content of expression are capable of altering the obligations of others. Someone can rationally think 'I want to run

[45] Ronald Dworkin, *Taking Rights Seriously* (Harvard 1977) xi.

that stop sign because I am late' but also say 'I am obligated to stop because it is the law'. Similarly, one can rationally think 'I want to break my agreement because it was a bad deal' and also say 'I am obligated to follow the agreement because it is a contract'. The features of the agreement that make it a bad deal (e.g. it will cost a lot of money) do not depend on the features which make the agreement a contract (e.g. there was valid consideration).

On the other hand, hate speech cannot alter the obligations of others because its force depends entirely on a normative claim which cannot be separated from the expression. All it can do is oblige the listener. The normativity of hate speech relies on the listener accepting the idea that one should be treated differently because of a protected characteristic. It makes no sense to say 'Adam should be treated differently because of his race' and also say 'people should not be treated differently because of their race'. The reason offered by hate speech to treat someone differently (the targeted person possesses a certain characteristic) has no force without accepting a claim *about the same thing*. One does not listen to hate speech and believe themselves constrained by an obligation to do a certain action despite one's wishes to not be bigot. Put another way, while there may be a group of bigots who recognise a discriminatory social hierarchy, nothing about hate speech obligates one to join this group.

4. Legal Implications

In this section, I discuss the doctrinal implications of my analysis for the legal regulation of hate speech. Specifically, I focus on three legal systems: that of the United States, the Council of Europe, and England and Wales. In what follows, I evaluate the

legal position of each in turn, with special reference to the differences in their approach.

Before doing so, however, I wish to discuss a feature shared by all three legal systems. Whether in America,[46] the Council of Europe,[47] or England,[48] the law does not generally consider the communication of justificatory reasons as a relevant factor by which to modulate the level of legal protection. Importantly, this serves as a decisive point of departure between these legal systems and my normative analysis.

To be clear, I am not suggesting that legal protections ought only to extend to speech which communicate justificatory reasons. In practice, however, the legal protections conferred on nonverbal speech often relate to the ability of a person to communicate the adoption or rejection of a certain group affiliation. This includes, for instance, the wearing of a badge or the burning of a flag. To my mind, protections regarding this sort of activity can plausibly be understood as parasitic to a right to freely associate with others, as distinct from the free speech principle. This is because the act of joining or leaving a group, or the continuing act of membership, is not distinctive to speech. So

---

[46] *Tinker v Des Moines* 393 US 503, 505 (1969) (extending equivalent free speech protections to the expressive act of wearing a black armband); *Texas v Johnson* 491 US 397, 404-06 (1989) (extending equivalent free speech protections to the expressive act of burning an American flag).

[47] *Fáber v Hungary* App no 40721/08 (ECtHR, 24 July 2012) at [36] (extending equivalent protections to the symbolic display of a flag); *Murat Vural v Turkey* App no 9540/07 (ECtHR, 21 October 2014) at [53] (stating that Article 10 protections apply equally to 'non-verbal and symbolic means of expression').

[48] *Monson v Tussauds* [1894] 1 QB 671 at 692 (libel typically takes the form of writing, but is equally capable of taking the form of 'a statute, a caricature, an effigy, chalk marks on a wall, or pictures'); 3 Bl Comm 125 (defining libel to include 'pictures, signs, and the like').

far as it involves expression, the act is best described as a speech-act, whose dominant constitutive element is its action component. As such, I argue that the protections conferred upon such speech-acts should follow from the protections afforded to a more general class of acts, namely those which relate to group membership.

While the legal position of hate speech does not currently depend on the communication of justificatory reasons, the concept is not entirely alien to the positive law. The most explicit adoption of this distinction, perhaps, lies in the American case law regarding the free speech protections of campaign contributions. In the leading case of *Buckley v Valeo*, the US Supreme Court made a crucial distinction between independent expenditures and campaign contributions.[49] The former consists of money used to promote a political message while the latter consists of money transferred directly to a political campaign's coffers. In *Buckley*, the Court held that the First Amendment right of free speech protects the ability of a person or group to spend money on independent expenditures, but not the ability to donate a certain quantum of money to political campaigns. Why? According to the Court, 'a contribution serves as a general expression of support for the candidate and his views, but does not communicate the underlying basis for the support'.[50] This, in my view, is best read as highlighting the distinctive value of justificatory reasons. While a campaign contribution simply signals support for a particular candidate, an independent expenditure is capable of publishing a message which communicates various reasons which attempt to justify a political position.

[49] *Buckley v Valeo* 421 US 1, 20 (1976).
[50] *ibid* at 21.

A. The United States

Here, I apply my account of the free speech principle to critically evaluate the legal position of hate speech in the United States. As we will see, the American approach is significantly more tolerant of hate speech than its contemporaries.

In 1992, the US Supreme Court handed down its decision in *RAV v City of St Paul*.[51] There, it struck down a Minnesota statute intended to punish cross-burning and other expressions of religious or racial animus. Controversy ensued.[52] After all, the American courts have long recognised the conditional nature of the right to free speech. For example, in *Chaplinsky v NH,* the US Supreme Court carved out a 'fighting words' exception to the First Amendment, defined as words 'which by their very utterance, inflict injury or tend to incite an immediate breach of the peace'.[53]

Enter the St Paul Bias-Motivated Crime Ordinance:

> 'Whoever places on public or private property a symbol, object, appellation, characterization or graffiti, including, but not limited to, a burning cross or Nazi swastika, which one knows or has reasonable grounds to know arouses anger, alarm or resentment in others on the basis of race, color, creed, religion or gender commits

[51] *RAV v City of St Paul*, 505 US 377 (1992).
[52] *RAV* has sustained vigorous criticism in the legal scholarship. For a semantic criticism, see M. Facchini and P.A. Grossman, 'Metaphor and Metonymy: An Analysis of R.A.V. v. City of St. Paul Minnesota' (1999) 12 *International Journal for the Semiotics of Law* 215-21. For a criticism on its constitutional analysis, see Andrea L. Crowley, 'R.A.V. v. City of St. Paul: How the Supreme Court Missed the Writing on the Wall' (1993) 4 *Boston College Law Review* 771-801.
[53] *Chaplinsky v New Hampshire*, 315 US 568, 572 (1942).

> disorderly conduct and shall be guilty of a misdemeanor.'[54]

The ordinance appears to share a similar rationale to the fighting words exception in *Chaplinsky*. Speech which arouses 'anger, alarm or resentment' is likely to fall within the fighting words exception. Indeed, the Minnesota Supreme Court interpreted the ordinance as only covering fighting words, and this interpretation was accepted by the US Supreme Court.[55] Yet, in *RAV v City of St Paul*, the Court found the ordinance unconstitutional. Scalia J, writing for the majority, held that the ordinance, while otherwise valid, placed restrictions on its application which were not required by its rationale.[56] In other words, an expressive act can be expressive for multiple reasons. A law which restricts expression cannot pick and choose which types of speech to proscribe, if it is justified on a rationale which would cover a broader class.

Thus, *RAV* held that a prohibition of threats, limited to those directed at certain protected classes, fail because regardless of how 'vicious or severe' the expression, it must fall within 'specified disfavored topics' for the proscription to apply.[57] The law limits its scope to expression that targets particular classifications. This, the US Supreme Court held, cannot stand, in part because the statute discriminates in determining which categorisations are impermissible and which are allowed. This discrimination, in turn, reveals a content-based rationale.

So far, so good. My account of the free speech principle chimes with the Court's reasoning. Both focus on the presence of a content-based rationale. Where they diverge, however, is in the

---

[54] St Paul Bias-Motivated Crime Ordinance, s292(2).
[55] *RAV* (n 51) at 381.
[56] *ibid* at 391.
[57] *ibid*.

Court's failure to distinguish between hateful conduct and the communication of reasons to hold a hateful belief. This failure is especially striking on the facts of *RAV*, which dealt with a defendant who burned a cross on the lawn of an African American's home. Such conduct, while expressive, provides no reasons.

The holding in *RAV* tracks the general position in American law. The strength of free speech protections does not usually turn on the communication of justificatory reasons.[58] It does, however, recognise a difference between pure[59] and symbolic[60] expression. As a descriptive matter, this broadly tracks the distinction between verbal and non-verbal speech. As a normative matter, however, the US approach diverges from my analysis. Rather than ground the distinction on the capacity of the means of expression to communicate reasons, as I do, it is premised on a division between the communicative and action component of symbolic acts. Consider, for instance, a person who vandalises a statute of a national leader. The act might communicate a message disapproving of the leader, but it also consists of the act of destroying property.

The American approach leads to a tripartite division. Pure speech benefits from the full protection of the First Amendment. Bare non-expressive acts lack any special protections. In the middle lie symbolic acts, which are only partially protected. The protection is partial because the speech and non-speech elements are analysed separately, with a

[58] One exception, as discussed above, lies in the legal protection of political contributions: text to n 49-50.
[59] *Cox v Louisiana*, 379 US 536, 555 (1965).
[60] *Thomas v Collins*, 323 US 516, 540 (1945).

compelling interest in regulating the latter justifying the incidental restriction of the speech element.[61]

This approach focuses on the reasons which a state has to regulate the action component of the symbolic act. I do not disagree with this part of the analysis. At the balancing stage, however, the courts appear to understand the value of the speech element of the symbolic act as equivalent to the value of protecting pure speech.[62] Given the well-established exception for time, place, and manner restrictions for even 'pure' speech,[63] it appears the only difference between pure speech and symbolic acts lies in the range of potential content-neutral justifications. That is, symbolic speech, since it encompasses a more diverse array of acts, is likely to implicate a wider range of content-neutral reasons to restrict its performance. Thus, as a matter of normative theory, I argue the American approach errs in treating the speech element of the symbolic act as having equivalent status to pure speech. The speech element of symbolic acts, so far as they do not rely on language, have a diminished capacity to supply justificatory reasons.

To be clear, however, *RAV* does not require the state to permit all acts of cross-burning. This can be seen in *Virginia v Black*, where the Supreme Court was able to affirm a statutory provision which bans cross-burning with the intent to intimidate.[64] To do so, the Court asked, following *RAV*, whether the statute adopted a content-dependent restriction on the scope of its applicability. Crucially, the only limitation on the statute's applicability was the presence of an intent to intimidate. Since it did not further limit its applicability to those intending to

---

[61] *United States v O'Brien*, 391 US 367, 376 (1968).
[62] *ibid* at 377.
[63] *Feiner v New York*, 340 US 315, 320-21 (1951).
[64] *Virginia v Black*, 538 US 343, 363 (2003).

intimidate a certain protected class, this statutory provision was upheld.[65]

B. The European Court of Human Rights

Most liberal states are far less tolerant of hate speech than the United States. They generally adopt a less capacious conception of expressive freedom. In doing so, they exclude from protection expression which incites hatred or offence.

Let us turn to the jurisprudence of the European Court of Human Rights. Consider the recent Chamber decision in *ES v Austria*:

> […] the exercise of the freedom of expression carries with it duties and responsibilities. Amongst them, in the context of religious beliefs, is the general requirement to ensure the peaceful enjoyment of the rights guaranteed under Article 9 to the holders of such beliefs including a duty to avoid as far as possible an expression that is, in regard to objects of veneration, gratuitously offensive to others and profane.[66]

This judgement must, of course, be considered in context: it is a supranational court deferring to the political process of Austria and its application of its own national laws. Hence, the ECtHR applies its margin of appreciation doctrine, wherein states are offered special deference when regulating expression in the spheres of morals or religion.

Its reasoning, however, runs contrary to the free speech principle. The ECtHR held that there is a legal duty to refrain

---

[65] *ibid* at 362. However, a plurality of the Court facially invalidated a statutory provision which states that cross-burning is prima facie evidence of an intent to intimidate: *ibid* at 367.

[66] *E.S. v Austria* App no 38450/12 (ECtHR, 25 October 2018) at [43].

from causing gratuitous offence. To do so, the Court applied its prior decision in *Otto-Preminger-Institut v Austria*, which found a right to 'proper respect for their freedom of thought, conscience, and religion'.[67] Then, in *ES v Austria*, the Court balanced this right with the separate right of the applicant to 'impart to the public her views on religious doctrine'.[68]

This decision is flatly inconsistent with my account of free speech. On the facts, the applicant held seminars where she discussed, in detail, her reasons for holding the Prophet Muhammad in disdain. This, inter alia, included her judgement that his marriage to Aisha constituted paedophilia. Because these inflammatory statements concerned the Prophet Muhammad, the Court held that they were likely to cause disruption in a way 'capable of hurting the feelings of the followers of that religion'.[69]

The facts of *ES v Austria* fall within the core case of protected speech. In an attempt to convince her audience, the applicant provided various reasons for her claim. The ECtHR, in delivering its judgement, endorsed the reasoning of the Austrian courts as it relates to the adequacy and quality of the reasons which the applicant gave. This sort of evaluative inquiry, however, runs contrary to the free speech principle. The principle, in preserving the ability to offer reasons, must protect the right to offer *bad* reasons for belief. To hold otherwise would entail the state arrogating for itself the right to judge its merits.

---

[67] *Otto-Preminger-Institut v Austria* App no 13470/87 (ECtHR, 14 January 1993) at [55].
[68] *E.S.* (n 66) at [46].
[69] *ibid* at [53].

C. England and Wales

Finally, let us turn to the position of hate speech in English law. The UK Supreme Court has described the right to political speech as 'a freedom of the highest authority'.[70] As Lord Steyn says, 'in a democratic society it is the primary right: without it an effective rule of law is not possible'.[71] Against the backdrop of the Convention,[72] the concept of 'political' speech has, in turn, been interpreted broadly.[73] Generally, it covers all speech that is expressed in the public sphere, on matters pertaining to topics of public interest.[74]

As such, the common law affords broad protections for political speech. The courts have affirmed that the free speech principle covers the 'irritating, the contentious, the eccentric, the heretical, the unwelcome, and the provocative'.[75] Thus, it protects speech which 'right-thinking people' consider dangerous or irresponsible, and speech which shocks and disturbs.[76] It does not, however, protect provocations of violence.[77]

These principles form the backdrop to an analysis of the primary hate speech offences in English law, which can be found in sections 18 and 19 of the Public Order Act 1986.

---

[70] R (*ProLife Alliance) v BBC* [2003] UKHL 23, [2004] 1 AC 185 at [6].
[71] *R v Secretary of State for the Home Department, ex parte Simms* [2000] 2 AC 115 at 125.
[72] *Thorgeirson v Iceland* (1992) 14 EHRR 843 at [64].
[73] Richard Clayton and Hugh Tomlinson, *The Law on Human Rights* (2nd edn, Oxford 2000) 15.284.
[74] *R (Calver) v Adjudication Panel for Wales* [2012] EWHC 1172 (Admin), [2013] PTSR 378 at [80]; Barendt, *Freedom of Speech* (n 27) 152.
[75] *Redmond-Bate v Director of Public Prosecutions* [2000] HRLR 249 at [20].
[76] *R v Central Independent Television plc* [1994] Fam 192 at 202.
[77] *Redmond-Bate* (n 75) at [20].

> **Section 18**
>
> A person who uses threatening, abusive or insulting words or behaviour, or displays any written material which is threatening, abusive or insulting, is guilty of an offence if—
>
> (a) he intends thereby to stir up racial hatred, or
>
> (b) having regard to all the circumstances racial hatred is likely to be stirred up thereby.[78]
>
> **Section 19**
>
> A person who publishes or distributes written material which is threatening, abusive or insulting is guilty of an offence if—
>
> (a) he intends thereby to stir up racial hatred, or
>
> (b) having regard to all the circumstances racial hatred is likely to be stirred up thereby.[79]

The broader statutory scheme of the 1986 Act points to a content-neutral rationale for these offences, namely in balancing the interest in expressive freedom with that of public order. This, I argue, is a less objectionable approach to the regulation of that speech. The proscription of speech likely to incite widespread disruption to public order is, in most circumstances, justified.

A speech regulation justified on the basis of protecting public safety would focus on the consequence, as opposed to the content, of speech. As such, the provision and consideration of reasons would not be directly targeted. Thus, in principle,

---

[78] Public Order Act 1986, s18(1). The offence was extended to cover religious hatred by the Racial and Religious Hatred Act 2006. It was further extended to cover hatred on the basis of sexual orientation by the Criminal Justice and Immigration Act 2008.
[79] Public Order Act 1986, s19(1).

individuals would be able to offer the entire gamut of reasons, so far as they took care to do so in a way unlikely to provoke disorder.

However, given the restrictions on its applicability, it is highly unlikely these offences have a content-neutral basis. Specifically, the offences only apply if the expression relates to hatred on the basis of certain protected characteristics. Further, the language of 'threatening, abusive, or insulting', absent an extraordinarily narrow interpretation, is too broad to plausibly support a public order rationale. Its wide scope means that the racial hatred requirement, in practice, serves as the crucial limiting principle for these offences.

This limits the range of reasons which a speaker is able to publicly express on the basis of racial hatred. The difficulty with this is that the racial hatred element of the offence is likely justified on the basis of a content-dependent rationale. In other words, it is justified by the wrongness of the underlying claim about race. This, however, would violate the free speech principle.

In response, it is tempting to take an empirical turn and argue that racial hatred is exceptionally likely to lead to a disruption of public order. This would reconcile the rationale of these racial hatred offences with the requirement for a content-neutral justification under the free speech principle. It also has the upside of fitting with the broader statutory context of the 1986 Act, which sets out a scheme which broadly aims for the protection of public order. There are, however, reasons to doubt this view.[80]

[80] Eric Heinze, 'Viewpoint Absolutism and Hate Speech' (2006) 64 *Modern Law Review* 577-78.

Beyond the empirical point, there are three points of tension between the current approach in English law and the public order rationale. Taken together, they throw serious doubt on the content-neutrality of these racial hatred offences, although they are not individually dispositive.

First, these offences, as currently interpreted, do not require 'proof that anybody actually read or heard the material'.[81] All that is needed is that the material is made 'available to the public',[82] regardless of the likelihood that others would come across it. For instance, an online post in an obscure blog would fall within this standard. This cuts against the rationale of public order. Speech which is never heard cannot cause unrest among the public. Of course, criminal offences often have a prophylactic element aimed at preventing harm before it happens. If the risk of disruption is low, however, the case for a prophylactic measure is weakened.

Second, these offences apply extraterritorially. The Court of Appeal has rejected the submission that the racial hatred at issue must be 'intended or is likely to be stirred up in this jurisdiction'.[83] In other words, an offence under section 19 may be sustained, even if the hatred 'may be stirred up overseas'.[84] However, the public order rationale is strained by the coverage of speech which only causes unrest in extraterritorial jurisdictions.

Third, there exists a parallel statutory regime which, in a content-neutral fashion, protects against disruptions of public order. It does so, admittedly, in a rather circuitous way. Specifically, section 89(2) of the Police Act 1996 makes it an

---

[81] *R v Sheppard and other* [2010] EWCA Crim 65, [2010] 1 WLR 2779 at [35].
[82] *ibid* at [34].
[83] *R v Lawrence Burns* [2017] EWCA Crim 1466, [2017] 9 WLUK 71 at [10].
[84] *ibid* at [11].

offence to willfully obstruct a police constable in the execution of his duty. Among the duties of a police constable is to prevent a breach of the peace. Thus, under the 1996 Act, constables have a duty to prevent the expression of speech likely to incite a breach of the peace, and a refusal by speakers to comply constitutes an offence under section 89(2). Importantly, this provision requires the existence of a 'real threat'[85] to public order. The scope of this restriction on freedom of speech is therefore limited to its consequentialist rationale. The presence of this parallel statutory regime cuts against the idea that sections 18 and 19 of the Public Order Act are needed to preserve public order. They are better understood as serving a more distinctive function, namely the content-based regulation of hate speech in English law.

Conclusion

The American constitutional tradition has had an outsized influence in this area. This risks a parochial turn in free speech theory. The subject matter of free speech theory – that is, the object of protection – is often defined in terms of the First Amendment.[86] Further, much of the literature adopts an interpretivist posture, limiting its relevance to those who accept that methodology.

Framing the debate in this way has other negative effects as well. At first glance, it makes attractive the search for a single unifying principle which justifies the First Amendment's empire.

---

[85] *Redmond-Bate* (n 75) at [18].

[86] The text of the First Amendment is as follows: 'Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances.'

This project, however, is hindered by the diversity of its subject matter. Specifically, one should allow for the possibility of a pluralistic and non-distinctive justification. By pluralistic I mean the acceptance of various grounds to uphold expressive freedom. And by non-distinctive I mean the justification of expressive freedom in terms which also justify the protection of other non-speech interests.

Indeed, if free speech theory is to justify the protection of speech, broadly construed, it may need to abandon the search for this unifying principle. It is doubtful, I suggest, for a singular principle to justify the protection of speech in both public and private contexts, expressed through both verbal and nonverbal means.

In this article, I adopted an approach which consciously tries to avoid these problems. First, I constructed an ideal theory of the free speech principle. Second, I took a comparative perspective to highlight its implications for the regulation of hate speech.

This approach reveals a distinctive reason to protect free speech: to preserve a process of public justification. To say this, however, is to embrace a limited vision of the free speech principle, one which accepts that it cannot protect all forms of expression. Strikingly, it would not cover private speech, like texts exchanged between friends. To my mind, however, the regulation of this type of speech is objectionable for reasons not distinctive to free speech. It might, for instance, constitute an invasion of privacy to an extent unacceptable in liberal society.

This approach also reveals the complexities underlying a trans-Atlantic comparison of hate speech regulations. It is all too tempting to chalk up any differences to that of 'wild-west

cowboys versus cheese-eating surrender monkeys'.[87] Put this way, the comparison is bipolar – it considers the relative merits of an approach which prioritises individual liberty as opposed to social equality. Such a comparison risks devolving into a conceptually unsatisfying stalemate, one which pits two incommensurable values against each other. Comparing both ends of this spectrum to an ideal theory of free speech, however, introduces a new layer of normative analysis. Both the American and European approach have flaws, and they are flawed in ways which have little to do with the battle between liberty and equality.

Let us return to the claim with which this article began. The free speech principle stands for the idea that the regulation of speech requires special justification, notwithstanding the other restrictions on state activity. That it supplies a distinctive reason to protect speech is especially important in the context of hate speech. I take it as common ground that abusively discriminatory conduct in the public sphere, *if it were not speech*, could and should be subject to legal regulation. If an exception is to be made, it must be justified on the basis of the free speech principle.

In this article, I have proposed one way to sketch the boundaries of this exception. While the state may regulate expression, it cannot intentionally target the expression of reasons which purportedly justify a claim. To do so requires the state to evaluate the content of those reasons. By taking sides on the basis of this evaluation, the state prevents its citizens from participating in a valuable process of public justification. It cannot do so while upholding its commitment to public reason.

[87] The phrase is borrowed from Eric Heinze, 'Wild-West Cowboys versus Cheese-Eating Surrender Monkeys' in I. Hare and J. Weinstein, *Extreme Speech and Democracy* (Oxford 2009) 182.