

Law as Counterspeech

Anjalee de Silva and Robert Mark Simpson

Ethical Theory and Moral Practice 26 (2023): 493-510

Abstract. A growing body of work in free speech theory is interested in the nature of counterspeech, i.e. speech that aims to counteract the effects of harmful speech. Counterspeech is usually defined in opposition to legal responses to harmful speech, which try to prevent such speech from occurring in the first place. In this paper we challenge this way of carving up the conceptual terrain. Instead, we argue that our main classificatory division, in theorising responses to harmful speech, should be between pro-discursive and anti-discursive responses. Some legal responses to harmful speech, so we argue, make a positive discursive contribution in their own right. That is, legal restrictions on harmful speech can have a function that is importantly similar to speech that aims at countering the effects of harmful speech.

1. Introduction

How should we combat the harms of hateful or vilifying speech? It seems natural to divide our responses into two categories. One option is speech-restrictive laws and regulations. The other option – one whose advantages and limitations have recently been explored by a number of authors – is *counterspeech*. In short, “hateful views can be countered by more public speech, which challenges hateful utterances” (Lepoutre 2017: 3; see also Langton 2017, Caponetto 2018, McGowan 2018, Lepoutre 2019a, Tirrell 2019, Howard 2020, Fumagalli 2021).

Why does it seem natural to categorise our options like this? First, for free speech hard-liners, who generally oppose speech-restrictive laws, it makes sense to separate acceptable and unacceptable responses to harmful speech. If censorship is unacceptable, you will naturally look for extra-legal ways of mitigating harm instead, and *counterspeech* is a way of naming that class of options. Second, regardless of whether one is a free speech hard-liner, it seems unfair for the burdens of redressing harmful speech to fall on the people targeted by it. Some recent work proposes counterspeech strategies that alleviate this unfairness (e.g. Howard 2020, Gelber 2021, Fumagalli 2021). But still, there is some

sense in distinguishing speech-restrictive laws, where the state bears the cost of dealing with harmful speech on everyone's behalf, from counterspeech, where the costs are more likely to be unfairly distributed.

We want to defend a different way of conceptualising the terrain. Insofar as counterspeech is essentially a matter of using more speech to counter harmful speech, it is possible to interpret some speech-restrictive laws themselves as a form of counterspeech. The law v. counterspeech classificatory framework tends to obscure the discursive potential of law. Our alternative way of dividing up the terrain, then, is to distinguish *pro-discursive* and *anti-discursive* responses to harmful speech. Pro-discursive responses are those which at least partly function to mitigate the effects of harmful speech, by communicatively contributing to the discursive milieu in which that speech occurs, in a way that counteracts the harm. Anti-discursive responses, by contrast, are those that simply aim to suppress harmful speech, without also (plausibly) making a positive communicative contribution to the discursive milieu in which that speech occurs.

One virtue of this framework is that it dispels the idea that a given response to harmful speech has a pro-discursive nature, purely by virtue of the fact that it is enacted via ordinary utterances, rather than a legal apparatus. Consider the person who yells *ad hominem* abuse at others who are engaged in harmful speech. This person may not be contributing in any meaningful way to the discursive milieu that elicits his response. In some cases, 'ordinary' counterspeech responses to hateful speech – that is, responses which are spoken or written, and which make no use of the law – can suppress other people's expression in ways that seem inimical to an ideal of open discussion.¹ The pro-discursive v. anti-discursive classificatory framework makes this salient, in a way that the law v. counterspeech framework doesn't.

However, the key thing that favours our proposed classification is that it allows for a more complete analysis of the pros and cons of legal restrictions on harmful speech. Such restrictions seem to evince an indifference to the potential of discourse itself, as a medium for redressing harmful communicative acts. We think this appearance is sometimes misleading. Some legal restrictions on speech have the potential to activate or facilitate positive discursive effects and may be endorsed for that very reason. Conversely, a general reluctance to restrict harmful speech may reflect an indifference to the pro-discursive effects that such restrictions can realise.

We are not denying that some speech-restrictive laws simply suppress disapproved viewpoints. But we don't believe that all such laws should be seen as solely or primarily anti-discursive in this way. Our suggestion is not that all speech-restrictive laws may be definitively characterised as either pro-discursive or anti-discursive. Rather, our suggestion is that some speech-restrictive laws are, when viewed with a degree of interpretative

¹ Mill makes a similar point. Populist tyranny is oppressive when it is enacted through the state. But things can be even worse when the majority imposes itself through extra-legal forms of coercion. "When society is itself the tyrant", Mill says, "its means of tyrannising are not restricted to the acts which it may do by the hands of its political functionaries. Society can and does execute its own mandates: and if it issues wrong mandates... it practises a social tyranny more formidable than many kinds of political oppression... it leaves fewer means of escape, penetrating much more deeply into the details of life, and enslaving the soul itself (1991: 8-9).

charity, better understood as constructive communicative contributions in their own right, relevantly adding to (and sometimes also facilitating) discourse. Our framework highlights this, and thereby allows for a better and more thorough analysis of the case for and against particular speech-restrictive laws and regulations.

We start in §2 by addressing an initial worry, related to the general viability of ‘expressive justifications’ for speech-restrictive laws. In §3 we explain the pro-discursive aims of counterspeech, as it is ordinarily conceived, and explain how some legal restrictions on speech achieve these purposes, such that it makes sense to recognise or classify them as pro-discursive responses. In §4 we address the key objection to our proposal. Even when restrictions on speech are a part of a conversation, aren’t they aimed at ending that conversation, in a way that is inherently anti-discursive? We finish in §5 by discussing the origins of the famous free speech saying, “the solution to bad speech is more speech”, in a way that reinforces our critique of a reductive law v. counterspeech framework.

2. Expressive justifications for speech-restrictive laws

Our thesis is that some speech-restrictive laws can be seen as at least partly pro-discursive responses to harmful speech. Given that thesis, it seems natural to suppose that the pro-discursive effects of such laws constitute some part of the justification for enacting them. Granted, it is possible to hold the two theses apart.

P1 A particular speech-restrictive law, L, has a pro-discursive function

does not entail

P2 L’s enactment is (pro tanto) justified, at least in part, by virtue of its pro-discursive function.

However, if P2 is false, then while P1 may well constitute an illuminating observation, it wouldn’t have much normative significance. The initial worry we need to address, then, is that there is a *prima facie* compelling argument, in the literature on counterspeech, to the effect that restrictions on speech cannot be justified by appealing to their pro-discursive function. This argument comes from Maxime Lepoutre, whose work has significantly contributed to the development of debate on this topic. The argument goes like this. Suppose we think the state has a duty to speak back against the views conveyed in hate speech and the like, in order to rebut them and mitigate their harmful effects. Even if we believe this, why shouldn’t we think it is sufficient, in terms of fulfilling this duty, for the state to engage in what Corey Brettschneider (2012) calls *democratic persuasion*? State officials can make speeches or publish statements, decrying discriminatory views and promoting an ethos of justice. Why isn’t that enough to counteract harmful speech? How can we justify anti-hate speech laws, by appeal to their discursive benefits, given that there are other ways for the state to bring about similar discursive benefits, without censoring people? As Lepoutre says

The expressive defense of legal bans construes them, roughly, as a kind of speech, which conveys a message. But that is what counterspeech is centrally designed to do.

So, the expressive defense of bans makes it difficult to understand why bans are needed. After all, if the function of hate speech laws can readily be performed without imposing sanctions on speech... then it seems we should forego such laws. (Lepoutre 2019b: 274-75)

Lepoutre says speech-restrictive laws need *distinctive* pro-discursive benefits, in order for them to garner support from an argument invoking this good. And he doubts that they have such a distinctive function. Democratic persuasion can do any discursive work that ‘law as counterspeech’ may do.

For example, you may think speech-restrictive laws provide a more intense rebuttal of bad views than democratic persuasion (e.g. Brown 2015: 263). But the state’s attempts at democratic persuasion could simply be made more vehement. Alternatively: if the state limits itself to democratic persuasion, maybe it sends mixed messages, insofar as its legal toleration of malign views conveys *de facto* approval of those views (see e.g. Matsuda 1989). However, on Lepoutre’s view, an authority’s toleration of x only signals support for x if the authority makes no clarifying statements about x. On that way of thinking, as long as the state uses democratic persuasion to condemn hateful views, it need not be seen as endorsing them (Lepoutre 2019b: 294).

Alternatively, perhaps law can accomplish distinctive discursive work through its subtlety or implicitness – i.e. the fact that it is always communicating its message in the background – which makes it more discursively efficacious than explicit democratic persuasion? For the sake of argument, let’s suppose that this is the best account of the distinctive discursive function of anti-hate speech laws.² Lepoutre’s worry is that if this is our main justification for legally restricting harmful speech, it will only have justificatory force to the extent that the restrictions succeed in deterring bad speech. Roughly, speech-restrictive laws can achieve a distinctively subtle or implicit discursive rejoinder to harmful speech, but only if those restrictions succeed in suppressing occurrences of such speech (Ibid: 291). But then if our speech-restrictive laws are succeeding on that front, an expressive justification for them becomes superfluous. They are justified because they suppress malign speech. Their pro-discursive function ends up being justificatorily redundant.

We find this argument unpersuasive, on three counts. First, legal restrictions on hateful speech have distinctive pro-discursive functions that Lepoutre’s analysis overlooks. Plausibly, there is at least one message that democratic persuasion cannot effectively convey, but which speech-restrictive laws can. They can say “this speech does such serious harm that we are justified in overriding our ordinary commitment to free speech in order to suppress it.”

Second, we are sceptical about Lepoutre’s claims regarding justificatory redundancy. As we explain below, we think restrictions can have positive discursive effects that obtain independently of any “suppressing” function. But suppose we grant for the sake of argument that the discursive benefits of such restrictions are only achieved when they suc-

² See Waldron (2012: 87ff.) for a version of this argument for anti-hate speech laws.

ceed in suppressing harmful speech. We don't think that makes these benefits redundant. By analogy, suppose an education policy improves learning outcomes, and also, as a side-benefit, increases students' self-esteem. It seems odd to insist that the policy's esteem-boosting effect is justificatorily redundant because its realisation causally depends on the learning effect. More generally, it seems implausible to think that the justificatory significance of two factors, in relation to some policy, can be inferred based on facts about the causal relation between them. But this sort of inference seems integral to Lepoutre's redundancy-related objection.

Third, Lepoutre's account downplays the ways in which law's omissions can function discursively. As we discuss further in §3, the state's monopoly on coercive power endows its legal stances, including its omissions, with a special weight. We agree that an authority's toleration of *x* signals support for *x* more clearly if the authority makes no clarifying statement about *x*. But even when there are clarifying statements, toleration can still signal support for *x*, or apathy towards *x*, in particular policy contexts. Consider the state that tolerates speech vilifying women, while simultaneously restricting speech that vilifies on the bases of race and religion – as occurs presently, in many jurisdictions. This state is plausibly signalling a degree of relative indifference to the vilification of women. And this seems to hold even if the state uses democratic persuasion to condemn misogyny. If one allows that the law says something about bigotry, whether it restricts it or tolerates it, then eschewing the *selective* toleration of harmful speech may be a distinctive function of speech-restrictive laws, and hence a discursive justification for them.³

In light of these considerations, we don't think Lepoutre's argument shows that the justification for enacting a speech-restrictive law cannot appeal in part to the pro-discursive function of that law. The justificatory relation between these two points isn't our main focus, in what follows. But we have hopefully said enough to dispel the initial worry. It isn't merely academic to defend the claim that speech-restrictive laws have a pro-discursive function. The view that one may naturally seek to defend, building on that claim – namely, that pro-discursive functions are some part of the justification for enacting such laws – is not ruled out in advance by a decisive objection.

3. The discursive effects of counterspeech and law

In this section we explain the discursive aim of counterspeech, as it is ordinarily understood, and how some legal restrictions on speech can achieve these aims, such that it can make sense, in principle, to see them as being relevantly similar to counterspeech. To make this part of our account more concrete, we focus on one genus of bad speech, namely, *vilifying speech*, i.e. speech that constitutes the discriminatory treatment of people in oppressed groups, and therein constitutes the subordination and silencing of these

³ Note that US constitutional doctrine seems to cast doubt on Lepoutre's reply to the 'mixed messages' worry. In *Brown v Board of Education* the courts found that segregation policies convey a derogatory view, irrespective of any government avowals of equality or disclamations of racism. For discussion of the implications of this vis-à-vis the legal toleration of hate speech, see e.g. Lawrence (1990).

people, on the basis of their relevant ascriptive characteristics (de Silva 2020).⁴ And within that genus, we are going to focus on one particular species of vilification, namely, *sex-based vilification*, i.e. speech vilifying women on the basis of their actual or perceived female sex. (Although our account should apply, *mutatis mutandis*, to other species of vilification.)

In examining the aims of counterspeech, we need an account of the harms that it is supposed to combat. We present a brief account in §3.1, drawing on parts of Rae Langton's and Mary Kate McGowan's linguistic pragmatics. The key theoretical commitment in this account, for our purposes, is that it analyses the harm of speech through the lens of social norms. One of the significant ways that ordinary speech effects harm is by re-enacting and reinforcing unjust social norms. And counterspeech aims to mitigate these harms either by blocking these effects or by enacting rival norms. Of course, this account is contestable. But the key theoretical commitment in it, as described, is one that is widely accepted among authors engaged in philosophical debates about counterspeech, and who are the principal audience for our argument. One can still follow our argument while plugging in a different account of the harms of hateful speech, as long as that commitment remains in play.

3.1 Constitutive and causal harms

Vilifying speech subordinates and silences its targets. Following Langton and McGowan, we posit that in systemically oppressive societies, vilification *constitutes* these harms, as opposed to just causing them. Sex-based vilification constitutes women in patriarchal societies as subordinated and silenced actors. What this means in practice, is that sex-based vilification enacts *de facto* social norms, the ongoing existence of which is a key part of what it means for a *de jure* egalitarian society to nevertheless be *de facto* sex-oppressive. We aren't saying that sex-based vilification creates sexist or misogynistic norms *ex nihilo*. Rather, it is a process of ongoing re-enactment. Sex-based vilification presupposes that women are already subordinate, and it does this in a context in which the relevant hierarchical norms are a determinant of women's social status. Thus sex-based vilification simultaneously relies on existing sex-based hierarchies, in order to attain its putative authority, while re-establishing those same hierarchies. Sex-based vilification ranks women as inferior on the basis of their sex, and thus legitimates women's inferior treatment. This ranking and legitimating has authority as a result of women's already-subordinate status, while also simultaneously (re)constituting women as subordinate. It also constitutes women as having nothing of worth to say, or no business speaking, and, in that sense, it silences them.⁵

⁴ The following section draws on work by de Silva (2020, 2021).

⁵ Another way that women are silenced by sex-based vilification, that Langton and others discuss, involves *illocutionary disablement*, i.e. undermining the conditions needed in order for women to perform certain acts they intend to perform via their speech, e.g. refusing sexual advances. While we don't deny that such silencing occurs, our analysis doesn't rely on this particular technical notion of silencing and applies just as much to more mundane forms of communicative suppression and inhibition.

In speech act theoretic terms, silencing and subordination are illocutionary, not just perlocutionary, results of sex-based vilification. In other words, these results aren't just caused by sexist and misogynistic speech, e.g. via processes of persuasion, influence, or conditioning. Rather, women are subordinated and silenced *in* being subject to sex-based vilification. In part this is because sex-based vilification dictates what moves women can make in the conversations where it occurs, via what McGowan (2009) calls *conversational exercitives*. But the constitutive harms don't end there, because those conversations don't exist in a vacuum. They are part of a continuous social fabric. In constituting women as subordinate in speech interactions, sex-based vilification constitutes women as subordinate in society *per se*. It reaffirms and carries forward the hierarchical rules and structures that place women below men (McGowan 2019).

We said that this isn't merely a result of persuasion or influence. But sex-based vilification has an impact on that level too. The illocutionary/constitutive harms of sex-based vilification are reinforced by perlocutionary/causal effects. "The beliefs of speakers and hearers", e.g. about how women are to be treated, "change in response to the abstract conversational score" (Langton 2012: 84). On Langton's account, this occurs through a process of *accommodation*, in Lewis's (1983) sense of the term. People's attitudes automatically adjust to be "whatever is needed to make sense of what is going on, thereby building up a ground of common belief" (Langton 2012: 73). For example, in a society where rape myths freely circulate, their tenets (e.g. that a woman's *no* often means *yes*) become widely accepted, as people's attitudes adjust into alignment with their discursive milieu. Preferences, desires, and emotions can similarly accommodate themselves to align with the relevant norms.⁶

In sum, sex-based subordination and silencing are two-sided processes. There is a re-enactment of society's sex-based norms, e.g. "women are to be treated as inferior". This is the crux of the illocutionary/constitutive harm. Simultaneously, people's attitudes are moulded towards compliance with those norms. This is the perlocutionary/causal harm. Sex-based vilification subordinates and silences in and of itself and it influences people's behaviour to align with, and to thereby fortify, this subordination and silencing.

Granted, there are empirically challenge-able aspects to these claims. In particular, there are important questions about the comparative importance of sex-based vilification, relative to other factors that foster sex-based inequalities. There are material factors – pertaining to employment, property, violence, and family life – that partly constitute women's unequal status. There are also other discursive ingredients, e.g. pop culture representations of women that aren't vilifying, but which plausibly still play a role in constituting sex-based hierarchy. To say sex-based vilification constitutes women's subordination isn't to say that it single-handedly constitutes that fact, just that it plays a

⁶ Sometimes the effects of sex-based vilification are worse than the mild language of 'accommodation' suggests. Vilified women may feel terrorised, and adapt their behaviours accordingly, e.g. by withdrawing from public spaces. Some acts of sex-based vilification are connected with intentionally terroristic forms of expression, like non-consensual pornography, see e.g. Citron and Franks (2014).

significant role in that fact's constitution. Similarly, to say that sex-based vilification perlocutionarily elicits compliance with norms of sex-based inequality, isn't to say nothing else exerts an influence in this direction.

3.2 *The aim of counterspeech*

Our picture is one in which the discursive ingredients of a systemically unjust society are neither causes, nor mere symptoms, of that reality, but rather, are coextensive with it. By analogy, if a couple is in the habit of spiteful quarrelling, the utterances through which their quarrels are played out are not mere symptoms of their conflict, but nor can they be construed simply as causes of it. Their conflict is partly constituted by the acrimonious communication that mediates the relationship. In a similar way, the subordinated position of women in a sex-oppressive society is constituted, in part, by the vilifying expression that articulates and gives manifest form to sex-based hierarchies.

The point of counterspeech, as that term is ordinarily used and interpreted, is to communicatively intervene in a way that counters the processes we have described. But how is this achieved? Sex-based vilification subordinates and silences by drawing on the already-operative power of sex-based hierarchies. This lends an implicit or *de facto* authority to sex-based vilification, even when it comes from the mouths of people who lack conventional forms of social authority (McGowan 2009). Sex-based vilification isn't creating an unjust social structure out of thin air. It is breathing life into an already-existing structure. Speech that aims to change the status quo cannot draw on the same sources of discursive power as speech which is sustaining the status quo. In order to counter sex-based vilification's subordination and silencing of women, therefore, counterspeech needs to draw upon some other kind of authority.

Such authority is difficult but not impossible to attain. Norms of sex-based inequality operate on a broad societal level, but there are rival norms operating alongside it at that level, and also in smaller cultural niches. Consider an analogy. Societies can be governed by predominant beauty norms, while also having space for rival beauty norms that apply in local regions or in countercultural niches. Or similarly: a country may have a persistent sub-culture of *de facto* tax avoidance, alongside *de jure* norms of tax compliance that are enforced rigidly in some regions, but loosely in others. So, while the norms of sex-based inequality operate in a society-wide manner, they aren't universally binding or immune to contestation. Speakers who contest the norms of sex-based inequality can enact norms in rival rule-governed activities, which contravene the wider norms of sex-based hierarchy (de Silva 2020: 1024).

This can occur in any number of ways. Feminist discourses in the media, in activism, in the arts, or in popular culture, can assert women's dignity and equal status. So too can universalistic ethical discourses, e.g. human rights theory. Given a sufficiently broad and positive reception, these forms of counterspeech can, in principle, illocutionarily constitute a nascent social reality in which women are not subordinated to men. They can also perlocutionarily encourage people towards practices that practically subvert women's silencing and subordination. This can range from modest gestures, e.g. people reading

more women authors, or men doing their fair share of the housework, to full-blown radicalism, e.g. the formation of women's separatist communities.

Nevertheless, non-state actors are limited in their power to shift social norms using counterspeech. Most people, influential celebrities and media magnates aside, will struggle to acquire a mass audience for their speech. And even the non-state counterspeaker who can reach a large audience may struggle to compel the assent of anyone who isn't already sympathetic to their view. Moreover, women who are silenced by sex-based vilification are greatly inhibited in speaking back against their own subordination. In a patriarchal society, vigorous criticism of sexism and misogyny can often be brushed aside. Consider an example from our home country. When the prominent Australian sports journalist Caroline Wilson spoke out against sex-based vilification directed against her by her colleagues in the media, she was roundly derided, and subjected to a torrent of misogynistic abuse – that is, further sex-based vilification – in online forums (Sherwood 2019). Even for more empowered speakers, who have a high-impact platform, speaking back against the norms enacted by sex-based vilification will often achieve patchy or even counterproductive results, in disrupting the predominant discourse and its corresponding norms.

We do not want to overstate the limits of what counterspeech can achieve. In particular, we would note that certain highly targeted instances of discriminatory speech, including sex-based vilification, probably cannot be effectively counteracted except by individual discursive contributions, which respond in a targeted way to the derogatory content or impact of the speech being countered. Token instances of counterspeech by non-state actors have some harm-mitigating power. But the contextual factors that empower sex-based vilification also impose some limits on the efficacy of such counterspeech.

3.3 *Law's efficacy, as counterspeech*

Now, suppose our society enacts a new anti-vilification statute, prohibiting certain forms of extreme misogynistic expression, e.g. including some forms of pornography and some misogynistic speech in online forums.⁷ The main intended impact of this, upon our society's discursive ecosystem, will be to reduce the incidence of some of the speech acts through which women's subordination and silencing are illocutionarily enacted and perlocutionarily reinforced. The threat of legal penalty deters people from carrying out acts of sex-based vilification, and this deterrence goes some way towards counteracting the perpetuation of harmful, sex-based inequalities.⁸ But as well as this, the speech acts involved in the law's enactment and administration can function as discursive rejoinders

⁷ We mean to be using an everyday vernacular definition, on which *pornography* just means sexually explicit representations aimed at sexually arousing. Some of the texts we are engaging with adopt a more theoretically-laden definition, inspired by Catharine MacKinnon, on which pornography is definitionally linked to women's subordination. That definition isn't built into our argument, though.

⁸ Although for an account of various doubts we might have about the efficacy such deterrence, see Stefancic and Delgado (1993).

to instances of sex-based vilification. This cluster of speech acts includes (i) various statements made in the process of the law's drafting and adoption, (ii) statements by judges and officials in the law's administration, affirming the law's anti-misogynistic purposes, and (iii) the legislative text itself. Through the combination of all these speech acts, the enactment and administration of this law can *say* – in an indirect manner, but nonetheless – something powerful to everyone in society. It can say

Yes, there are *de facto* social norms in effect in our society, according to which women being treated as second-class citizens is a perfectly normal and acceptable way for things to go. But these *de facto* norms are normatively invalid. Acts and customs that accord with them, that is, which treat women as inferior, and as unworthy of being listened to, are simply unacceptable.

In this discursive contribution, the law mitigates the subordinating and silencing effects of sex-based vilification, not just for men treating women as inferiors, but also for women accommodating their attitudes and practices to such treatment. A young woman who watches porn online, attempting to thereby figure out the norms of heterosexual sex, finds many depictions of men abusing women, and of women inviting and enjoying that abuse. She may be savvy enough to recognise that something is less than fully truthful in these depictions. But even so, a part of her ends up thinking that in the domain of sex, women (ought to) naturally desire male abuse. Our new law tells her otherwise and repudiates the hierarchical *de facto* norms that this idea serves. In place of the norm “women are to be treated as inferior”, it espouses a rival norm: “women must not be treated as inferior”. And it influences people towards acts that accord with the new, opposing norm. The enactment and administration of the law thus helps to achieve the desired discursive results of counterspeech: it counteracts the illocutionary/constitutive subordinating power, and the perlocutionary/causal subordinating influence, of sex-based vilification.⁹

The formal implementation of the state's *de jure* authority endows the law with a greater capacity to achieve these results than counterspeech by non-state actors. If a celebrity, journalist, or academic says “*de facto* norms of sex-based hierarchy are invalid”, a natural reaction for an audience member may be to think “but who are you to say so?” This scepticism isn't just a matter of doubting the counterspeaker's credibility. The audience member may think “who are you to say so?”, even if they see the counterspeaker as properly authoritative. The scepticism owes to the fact that *de facto* norms cannot be spoken out of existence so easily. That isn't how social norms operate. If person A says something that aligns with and re-enacts a society's unjust norms, he isn't simply shifting a discursive lever that person B can then shift back again via her own speech. Person A speaks with the weight of a whole social structure behind him.

Granted, if we can get a coordinated and influential network of people speaking back effectively against some entrenched social norm, that may be enough to effect the normative shift that's being sought. Our point here isn't to deny the norm-shifting potential

⁹ Gelber and McNamara (2016) discuss these kinds of discursive effects, and evidence of them in particular contexts. They draw on empirical studies to illuminate the messages of solidarity that the targets of racist vilification take themselves to be receiving through the enactment of anti-vilification laws.

of counterspeech altogether. Our point, rather, is about the relative efficacy with which law can alter social norms, given the authority and esteem that underwrites it. When the law is speaking back against speech that enacts oppressive social norms, it speaks both on behalf of society – insofar as it is the speech of a representative government – and in a way that is backed by the state’s monopolisation of the use of coercive power in a society. Naturally, this doesn’t guarantee that any particular instance of ‘law as counterspeech’ will achieve our desired discursive results. But it confers a marked advantage on its potential to achieve those results.

The question arises whether democratic persuasion performed by the state is just as (or more) effective in the relevant ways as law as counterspeech. If so, democratic persuasion, being in all ways non-speech-restrictive counterspeech, is arguably a more desirable way to combat the evils of bad speech in liberal democracies, than is law as counterspeech. The relative efficacy of democratic persuasion as compared to law as counterspeech is ultimately an empirical question, for which we cannot have a conclusive answer. Nevertheless, it is *prima facie* reasonable to suggest that the law, backed as it is by the formal implementation of the state’s coercive authority, is equally or more *expressive* of authority, and thus likely to be more effective than ‘mere’ democratic persuasion.

Moreover, to argue that democratic persuasion performed by the state could be just as effective as law as counterspeech, is to overlook the discursive functions of the law’s *omissions*, along the lines of our brief discussion above. If the law says something when it speaks *and when it doesn’t*, then avoiding the *selective* toleration of harmful speech is a function of speech-restrictive laws that is distinct from the functions of democratic persuasion. It is hence a reason to think that such laws, in certain contexts, at least, offer a more effective form of state counterspeech than democratic persuasion.

3.4 *Conceptual limits to counterspeech*

One may see all of this as being beside the point, though. Even if law is better placed to achieve the pro-discursive effects that counterspeech, including democratic persuasion, is aiming at, that doesn’t mean law *is* counterspeech. By analogy, law can advance the same aims as political activism, like policy reform. But that doesn’t mean that the law *is* – can somehow be equated with – political activism. Indeed, one may think that concepts like *political activism* and *counterspeech* ought to be defined through an explicit contrast with law. Whereas law relies on coercive force to effect change, counterspeech and political activism effect change via softer forms of power: persuasion, influence, consciousness-raising, or solidarity-building. Law is fundamentally characterised, one might argue, by its being a vehicle for coercion or compulsion. Counterspeech is supposed to be a way of counteracting harmful speech through other means.

There is nothing to stop one from defining counterspeech in a way that puts law outside of the concept’s scope by stipulation. But we are doubtful about the theoretical motivation for doing this. It is true that law as counterspeech implicates the use of coercive force, in a way that democratic persuasion by state officials doesn’t. And it follows from

this that law as counterspeech has to satisfy a higher justificatory standard than democratic persuasion, consideration of which standard includes, but is not limited to, the potentially greater efficacy of law as counterspeech. Policies that involve the use of coercive force answer to more stringent justificatory demands than ones that don't, *ceteris paribus*. However, this doesn't entail that there is anything benighted about classifying legal restrictions as counterspeech, at least in some cases. The aptness of that classification owes to the fact that some speech-restrictive laws make *bona fide* discursive contributions. And the aptness of that classification isn't nullified purely on account of the fact that those laws also involve coercion, and therefore must answer to more stringent justificatory standards. All that follows from this is that some species of counterspeech stand in need of justification more so than others.

Here is an example, to illustrate. A teacher wants to stop her students teasing each other. In scenario A she does this via an authoritative declaration of behavioural standards: "in this class we don't tease people." In scenario B she backs this up with a threat: "in this class we don't tease people unless we want detention." The similarities between the statements in A and B strike us as more important, in terms of how we classify various discursive actions, than the differences. Both statements involve an authority setting standards of conduct for her charges. The B-statement indicates that non-compliance comes with enforceable costs, where the A-statement doesn't. But that difference doesn't nullify the B-statement's discursive character. The two statements are doing similar things, in communicating standards of conduct, even if only one of them is simultaneously being backed up by coercive power.

The key point is that bearing some relation to coercion doesn't necessarily transform the discursive character of a communicative act. And thus, an utterance that can be partly classified under some discursive-classificatory concept, X, shouldn't automatically be seen as falling outside of X, just because it implicates coercive force. Again, one is free to stipulatively define *counterspeech* to exclude law as counterspeech, because of its coercive character. But we are arguing that a natural theoretical motivation that one may invoke, to justify this piece of conceptual housekeeping, isn't compelling.

In sum, then, law and counterspeech don't have to be seen as strictly opposed ways of responding to harmful speech. We have argued that anti-vilification laws can have a positive discursive function. The fact that they are backed by state authority makes it easier for them to reverse the tide of *de facto* norms that empower sex-based vilification, and which sex-based vilification perpetuates in turn. We can redraw the boundaries of the concept of counterspeech, therefore, in a way that allows some law to count as counterspeech. Plausibly, anti-vilification laws can achieve the same sort of pro-discursive effects that 'regular' counterspeech aims at, and they can do so more effectively.

4. Pro-discursive speech-restrictive laws

The law v. counterspeech classificatory framework tends to obscure the pro-discursive character of (some) speech-restrictive laws. The alternative way of dividing up the terrain, then, which we now want to defend, is to distinguish pro-discursive responses to harmful speech from anti-discursive responses. Pro-discursive responses – which include some instances of speech-restrictive laws and regulations – function to mitigate the bad effects of harmful speech, by communicatively contributing to the discourse in which that speech occurs, in a way that counteracts the relevant harms. This is not to say that all pro-discursive responses are paradigmatic instances of counterspeech, as it is ordinarily understood, but merely that responses may be classified as pro-discursive if they at least partly function in this way. Anti-discursive responses, by contrast, aim to suppress harmful speech, but they don't make a positive communicative contribution to the discourse in which that speech occurs.

4.1 *The ropes and the boxing match*

There is room for doubt, we concede, about whether the law says things, in a sense of *saying* that corresponds with the notion of a positive discursive contribution. A country's tax code says things about income tax rates, or deductible business expenses, but it isn't aptly construed as a contribution to public discourse about taxation. Rather, it implements policies that have been arrived at under the influence of that discourse. It isn't really a part of the conversation; it's just what the conversation is about. And something similar goes, so one might argue, for anti-vilification laws and the like. These laws restrict and guide our conduct. Whereas speech ordinarily achieves its aims through processes of persuasion or suggestion, legal restrictions *insist*. In general, law interacts with discourse by setting constraints around it. It is not aptly construed as a contribution to public discourse, any more than the ropes around the boxing ring can be construed as part of a boxing match.

To address this challenge, we first want to explicitly renounce any across-the-board thesis about the pro-discursive character of speech-restrictive laws. Different speech-restrictive laws call for different analyses, vis-à-vis their discursive nature. Some are rightly understood as mere constraints on public discourse, like the ropes around the boxing ring. Consider *lèse-majesté* laws, of the kind that apply in Thailand and Cambodia, which impose criminal penalties upon those found to have insulted the royal regime. It would be stretching any concept of *the pro-discursive* beyond credibility, if we were to claim that such laws contribute to – as opposed to simply constraining – public discourse. These laws have some communicative effect, as all laws do. But their function is so clearly about quashing discourse, that to characterise them as pro-discursive would seem absurd.

Our point, though, is simply that some speech-restrictive laws are unlike *lèse-majesté* laws, in this respect. The speech acts involved in the enactment and administration of some speech-restrictive laws *can* be aptly construed as contributions to discourse, and not merely as constraints upon it.

Which laws, and which communicative acts involved in their enactment and administration, are aptly seen as pro-discursive? First, some speech-restrictive laws include prefatory remarks that explain their purposes. When these remarks are read – by officials, researchers, political actors, or members of the public – they convey viewpoints on the issues they address, and thus contribute to public discourse in a way that isn't totally unlike the viewpoints conveyed in other pieces of writing on those issues. Second, consider the statements that comprise judicial rulings, in the administration of speech-restrictive laws. Such statements can express condemnations of wrongful acts that breach statutory law. Some involve the interpretation of constitutional provisions, or the establishment of common law precedents, which in either case can serve the dual function of specifying the scope of restrictions on speech and explaining the bases for doing so. Judicial statements may be addressed to a small proximate audience, but they can then be relayed via the media to wider audiences. Such statements can thus enter the broader public discussion about the issues to which the speech-restrictive laws pertain.¹⁰

What about legislative texts as such? Whether these constitute contributions to public discourse depends on various factors, including a law's prominence and expressive clarity. Crucially, speech-restrictive laws are more likely to constitute discursive contributions in their own right, if they are enacted in the midst of public debate around the issues they address. Consider the UK's 2006 Racial and Religious Hatred Act. This modified the 1986 Public Order Act to make religious vilification an offense on similar terms to racial vilification, per the Act's original formulation. The 2006 Act was reacting to a heightening of inter-faith tensions in the wake of the September 11 attacks. It was a result of the then-Labour Government's third attempt to pass anti-religious vilification laws, with earlier attempts having been blocked by Parliament. Given these contextual factors, the passage of this legislation can be seen as actively contributing to UK debates around toleration and religious identity. It functioned as a statement, by the state, on behalf of society, about the importance of respecting the equal status of religious minorities.

Not all speech-restrictive laws contribute to public discourse like this, such that they can be viewed as conveying a perspective on a current debate. Some anti-vilification laws may be enacted in a way that is more like 'legal housekeeping' than a conversational move, e.g. when they are occasionally enacted without debate or controversy, to fulfil states' duties under international treaties, like the duties in Article 4 of the 1969 Convention on the Elimination of All Forms of Racial Discrimination or Article 20(2) of the 1966 International Covenant on Civil and Political Rights.

¹⁰ One possible objection to our claims here, is that these forms of law-adjacent speech aren't *dialogical*, which arguably bars them from being seen as pro-discursive. Contributions to discourse, you might argue, must involve dialogue, like in a face-to-face conversation, or a written back-and-forth online. But this isn't how law speaks. Speech acts in law are more like one-way edicts, than contributions to a discussion. However, we don't see this as serious problem for our account. First, it just isn't clear why contributions to discourse must be dialogical. Reclusive authors publishing books, and neglecting to correspond with others, are surely still contributing to discourse, if their works are being read. Second, and relatedly, if this criterion were applied generally, it would mean that many of the communicative responses to harmful speech that are standardly classified as counterspeech (e.g. non-dialogical publications) will turn out to be non-discursive in character. And this seems implausible. We don't see any strong case for positing a fundamental conceptual link between the discursive and the dialogical.

In any case, we are not trying to mark out an artificially bright line. Accordingly, it is not in our view pivotal that perhaps only a small number of laws enacted in democratic societies may be ‘more’ anti-discursive than they are pro-discursive. We will discuss this further in §4.2. For now, we have defined pro-discursive responses as those which, given their context, function to mitigate the bad effects of harmful speech, by communicatively contributing to the discourse in which that speech occurs, in a way that counteracts the harm. Our aim here has been to indicate some of the forms that that discursive contribution can assume. When speech-restrictive laws are enacted they can be responsive, to a greater or lesser degree, to current debates in the society where they apply. Legislative texts that are enacted in response to such debates, and which are of interest and significance to many people, can be seen as making contributions to discourse, as opposed to merely setting constraints upon it.¹¹ That is, not all speech-restrictive laws are solely or primarily anti-discursive, and a pro-discursive v. anti-discursive framing illuminates and accounts for this in ways that that a more reductive law v. counterspeech framing does not.

4.2 *Having the final word*

But what if it turns out that all speech-restrictive laws have a practical impact on public discourse similar to *lèse-majesté* laws? We are claiming that the UK’s 2006 Racial and Religious Hatred Act can be seen as making a contribution to public debates about religious equality, given the discursive context in which it was enacted. But if the actual impact of the Act was to quash all discussion of Islam, or debate about religious equality, similar to the way that Thailand’s *lèse-majesté* laws quash criticism of the royal regime, then our overall thesis – that speech-restrictive laws like this can be understood as pro-discursive responses to harmful speech – still seems dubious. Whether a legal response to speech is pro-discursive or anti-discursive surely doesn’t just depend on whether it is in some conceivable sense making a contentful contribution to a discussion. It also depends on whether it allows that discussion to carry on, and on the discursive virtues or vices that it elicits therein. A contribution to a discussion that attempts to bring that discussion to a permanent finish seems pro-discursive only in a trivial sense (or indeed, an Orwellian one).

¹¹ Our account might be seen as being tacitly committed to an expressive theory of law. Some expressive theories emphasise the law’s function in coordinating activity by conveying information, e.g. about social attitudes towards certain acts (McAdams 2015). Others focus on how the law gives expression to values (Anderson and Pildes 2000). An expressive theory of law might thus interpret an anti-vilification statute as expressing a commitment to equality, or as conveying information about a society’s egalitarian ethos. The common feature in all expressive theories of law is a rejection of the thesis defended by certain authors (e.g. Schauer 2015), that the fundamental way law impacts our behaviour is via coercion, either active or latent. Expressive theories allow that law is underpinned by coercive power, but they say law’s impact on our lives isn’t defined by or reducible to that. Is our argument committed to this kind of theory? No. It is agnostic about them. Even if you think law is essentially about coercion, you can allow that the speech involved in law’s enactment enters public discourse. Critics of expressive theories needn’t deny that law generates discursive material. All they deny is that this material is of primary importance in our understanding of how law guides human activity. Granted, if you wanted to say that literally all speech-restrictive laws are pro-discursive function, this might commit you to an expressive theory of law. But that isn’t our position. We are saying some speech involved in the enactment of some speech-restrictive laws contributes to public discourse. We are denying that such laws can only ever be ‘the ropes around the ring’ of discussion.

We have three points to make in response to this. First, there are clearly significant differences in how different speech-restrictive laws impact on the discussions they are regulating. Whereas criticism of the monarchy in Thailand is effectively suppressed, vigorous criticism of Islam remains a feature of UK society. The effect of the 2006 Racial and Religious Hatred Act was to modify how such criticism is expressed, by deterring some of its more vitriolic manifestations. The Act contributed to discussions in the UK in which Islamophobic ideas were circulating, in a way that aimed at changing their complexion and mitigating certain discriminatory harms. It didn't seek to bring those discussions to a permanent conclusion. Free speech hard-liners may argue that changing the complexion of a discussion is still an incursion upon people's expressive rights. But there is room for reasonable disagreement on that front. And in any case, our point here is just to deny that all speech-restrictive laws try to bring debate to a close, *à la* Thai *lèse-majesté* laws.

Second, while it isn't necessarily misguided to worry about the slippery-slope potential of speech-restrictive laws that aim to make such discursive interventions, this is orthogonal to the classificatory point we're addressing. In practice, some restrictions suppress discussion entirely, while others just try to alter the complexion or orientation of a discussion, in a harm-mitigating way. The latter may be criticisable because of the slippery slopes they open up. But unless we are sliding down these slopes, it seems descriptively misleading to equate them with speech restrictions that aim to quash debate outright.

Third, it behoves us to consider the ways in which speech-restrictive laws can improve the discussions they regulate, not only by removing ignorant or vitriolic contributions, but also by opening them up to more participants. We can return to our example from §3 to illustrate. The vilification of women, especially in online media and social media contexts, deters women's discursive participation in those arenas. The idea of shutting some people up so that others can speak – what Owen Fiss (1991: 2059) calls a *parliamentarian* approach to free speech – seems apt in these cases. Of course, that idea can be invoked in bad faith, by authorities seeking to suppress views that they ideologically oppose. But one would have to be ultra-cynical to believe that this is what is occurring with all speech-restrictive laws that are defended by appeal to this idea. It is possible for some speech-restrictive laws to simultaneously act as discursive rejoinders to the subordinating effects of sex-based vilification, and to improve public discourse by making particular discursive domains less exclusionary spaces for particular groups. Restrictions on online sex-based vilification are one example of this. Some speech-restrictive laws don't just allow discussion to continue, they also enrich its discursive virtues, in particular, the diversity of the viewpoints that it accommodates.

Our response to this challenge leaves us in a position to summarise the features of the speech-restrictive laws that qualify as pro-discursive responses to harmful speech. At a general level of description, they are laws that aim to deter or mitigate the bad effects of harmful speech, while communicatively contributing to the discourse where that speech occurs. Their communicative contribution can come via legislative texts themselves, or via speech acts performed in laws' enactment or administration. The main feature that marks them out as making a positive contribution to discourse is that they are contextually responsive to active debates in the society where they apply. And finally, they do not 'contribute' to those debates in a way that simultaneously aims at bringing debate to a

close. In some cases, quite the contrary, their function is to positively enrich and edify the discourses to which they contribute.

These criteria – for what makes a speech-restrictive law count as a pro-discursive response to harmful speech – require discretion and interpretation in their application. They do not leave us with a mechanical sorting algorithm that can uncontroversially divide pro-discursive responses from anti-discursive ones. But they are a useful set of heuristics for mapping this intuitively plausible distinction onto real-world cases. Some speech-restrictive laws really are, purposively and functionally, just attempts to silence those who hold disapproved views. Others function to contribute to discussions and therein mitigate some of the harms effected by speech within them, as well as orient the discourse in a better direction. The account we have presented helps us to assess which is which. And to be clear, the fact that a given speech-restrictive law or regulation, L, is adjudged to be a pro-discursive response to harmful speech, doesn't by itself entail that L is justified, on the balance of considerations. But this judgement is a *pro tanto* justificatory consideration in L's favour. Our account's payoff is in (i) calling our attention to this *pro tanto* consideration, and (ii) providing guidance for the judgements involved in it.

5. Conclusion: the solution to bad speech

When classifying different ways of responding to hateful speech, and contemplating the pros and cons of different responses, we should centre the distinction between pro-discursive and anti-discursive responses, in place of a law v. counterspeech classificatory framework. The latter framework naturally invites us to suppose that the best way to promote healthy public discourse is by eschewing legal interventions and speaking back in other ways. But some speech restrictions provide meaningful contributions to the debates that they are purporting to regulate. Legal intervention is sometimes just as good a way of dealing with harmful speech, even if we are exclusively focusing on the question of what will make for the richest and most vibrant discussion.

If this position appears somewhat iconoclastic, this is partly because of how deeply the law v. counterspeech classification orients our thinking on these issues. Some responsibility for this can be attributed to the famous saying in work on free speech, that “the solution to bad speech is more speech.” This saying has its origins in a dissenting opinion from Justice Louis Brandeis, in the landmark US Supreme Court case *Whitney v. California* (1927), a case that concerned the government's right to quash seditious dissent. The familiar, short-hand version of this saying doesn't capture the nuances of Brandeis's thought, or the issue that he was addressing. What Brandeis said was

No danger flowing from speech can be deemed clear and present, unless the incidence of the evil apprehended is so imminent that it may befall before there is opportunity for full discussion. If there be time to expose through discussion the falsehood

and fallacies, to avert the evil by the process of education, the remedy to be applied is more speech, not enforced silence.¹²

Brandeis isn't saying more speech is *the* solution to bad speech. He is saying that it is *a* solution to some bad speech. And it is worth noting the type of bad speech, and the putative evil to be averted, that he has in mind, in these remarks. Charlotte Anita Whitney was a political activist from a wealthy, well-connected family. In 1920 she was convicted of criminal syndicalism by a California state court, over a talk she gave to a civic club promoting the newly-formed Communist Labour Party. Whitney's conviction was, to all practical purposes, a state-backed suppression of political dissent. It was upheld by the Supreme Court – whose free speech doctrine in the 1920s was yet to assume the robust liberal guise that it later took on – on the grounds that her speech had a dangerous tendency to destabilise government.¹³ The reasoning behind this ruling wasn't unlike the rationales we see authoritarian states using today, to justify the persecution of dissidents. Brandeis was condemning this, and pressing a rival, anti-authoritarian principle: you can't say someone is endangering society, and censor her speech on that basis, if she is just trying to persuade people that society should be changed. Brandeis wasn't endorsing the sweeping claim that today's proponents of the "more speech" slogan often make: that counterspeech is the only justifiable or effective way to combat malign speech. He was saying that serious dissent must be met with critical discussion, and hence that it is wrong to justify the punishment of dissidents by equating the fact of their dissent with a life-or-death political crisis.

The libertarian lesson that some authors take from Brandeis is based on a contestable analysis of his opinion. But at the same time, Brandeis surely was promoting the kind of law v. counterspeech classificatory framework that we have been criticising. He says the remedy for bad speech is more speech, as opposed to enforced silence. This rhetorically equates legal restriction with a silencing of discussion. It obscures the possibility that legal restrictions may themselves add to the *full discussion* Brandeis wants us to be having.

In revisiting Brandeis's saying, and more generally, in challenging the law v. counterspeech classificatory framework, we want to subvert the notion that hard-core libertarians are the true allies of discussion, and that anyone who favours moderate restrictions on hateful speech prizes discussion less than other things, like social equality. Proponents of sensible speech-restrictive laws may, and many of us do, aspire to a full and frank discussion, just as much as the free speech hard-liners who draw inspiration from Brandeis. Where we non-hard-liners disagree with the hard-liners, is in our view about the role that speech-restrictive laws can play in contributing to, and thus realising, this

¹² *Whitney v. California*, 274 U.S. 357 (1927).

¹³ In *Brandenburg v. Ohio*, 395 U.S. 444 (1969), the Supreme Court updated its *clear and present danger* doctrine (which specified the limits of permissible political dissent at the time of *Whitney*) with a doctrine on which restrictions on political dissent were only permitted for speech *intended* to, and likely to, incite *imminent* lawless action. Brandeis's opinion was one of a series of opinions, in the interwar period, adverting to the ways in which the clear and present danger test allowed for the illiberal suppression of dissident views. It thus pushed in favour of *Brandenburg's* eventual doctrinal reforms.

full and frank discussion.¹⁴ More speech is a good thing, but sometimes that very factor speaks in favour of enacting judicious restrictions on vilifying speech.¹⁵

References

- Anderson, Elizabeth S., and Richard H. Pildes (1996), "Expressive theories of law: a general restatement", *University of Pennsylvania Law Review* 148 (5): 1503-75.
- Brettschneider, Corey (2012), *When the State Speaks, What Should it Say? How Democracies Can Protect Expression and Promote Democracy* (Princeton: Princeton University Press).
- Brown, Alexander (2015), *Hate Speech Law: A Philosophical Examination* (Abingdon: Routledge).
- Caponetto, Laura (2018), "Undoing things with words", *Synthese* 197 (6): 2399-414.
- Citron, Danielle Keats, and Mary Anne Franks (2014), "Criminalizing revenge porn", *Wake Forest Law Review* 49 (2): 345-92.
- de Silva, Anjalee (2020), "Addressing the vilification of women: a functional theory of harm and implications for law", *Melbourne University Law Review* 43 (3): 987-1032.
- _____ (2021), *Addressing the Vilification of Women: A Functional Theory of Harm and Implications for Law* (The University of Melbourne, Doctoral Thesis).
- Fiss, Owen M. (1991), "Freedom and feminism", *Georgetown Law Journal* 8 (6): 2041-62.
- Fumagalli, Corrado (2021), "Counterspeech and ordinary citizens: how? when?", forthcoming in *Political Theory*.
- Gelber, Katharine (2021), "Speaking back" in Adrienne Stone and Frederick Schauer (Eds), *The Oxford Handbook of Freedom of Speech* (Oxford: Oxford University Press).
- Gelber, Katharine, and Luke McNamara (2016), "Anti-vilification laws and public racism in Australia: mapping the gaps between the harms occasioned and the remedies provided", *University of New South Wales Law Journal* 39 (2): 488-511.
- Howard, Jeffrey W. (2021), "Terror, hate and the demands of counter-speech", *British Journal of Political Science* 51 (3).
- Langton, Rae (2012), "Beyond belief: pragmatics in hate speech and pornography" in Ishani Maitra and Mary Kate McGowan (Eds), *Speech and Harm: Controversies over Free Speech* (Oxford: Oxford University Press).
- _____ (2017), "Blocking as counter-speech" in Daniel Fogal, Daniel Harris, and Matthew Moss (Eds), *New Work on Speech Acts* (Oxford: Oxford University Press).
- Lawrence, Charles R. (1990), "If he hollers let him go: regulating racist speech on campus", *Duke Law Journal* 1990 (3): 431-83.
- Lepoutre, Maxime (2017), "Hate speech in public discourse: a pessimistic defense of

¹⁴ Our point here owes a debt to one of MacKinnon's well-known remarks, in her arguments for restricting pornography. "The free speech of men silences the free speech of women", she says; "it is the same social goal, just other people" (MacKinnon 1983: 337).

¹⁵ For helpful discussion of these ideas and/or critical comments on earlier drafts of this paper, we are grateful to two anonymous referees from this journal, as well as Corrado Fumagalli, Jeff Howard, Enes Kulenović, Maxime Lepoutre, and the participants in an online workshop on counterspeech organised by Corrado Fumagalli in July 2021.

- counterspeech”, *Social Theory and Practice* 43 (4): 851-83.
- _____ (2019a), “Can ‘more speech’ counter ignorant speech?”, *Journal of Ethics and Social Philosophy* 16 (3).
- _____ (2019b), “Hate speech laws: expressive power is not the answer”, *Legal Theory* 25 (4).
- Lewis, David (1983), “Scorekeeping in a language game” in *Philosophical Papers, Volume I* (Oxford: Oxford University Press).
- MacKinnon, Catharine A. (1983), “Not a Moral Issue”, *Yale Law & Policy Review* 2 (2): 321-45.
- Matsuda, Mari (1989), “Public response to racist speech: considering the victim’s story”, *Michigan Law Review* 87 (8): 2320-81.
- McAdams, Richard H. (2015), *The Expressive Powers of Law* (Cambridge MA: Harvard University Press).
- McGowan, Mary Kate (2009), “Oppressive speech”, *Australasian Journal of Philosophy* 87 (3): 389-407.
- _____ (2018), “The more speech response, counter speech, and the complexity of language use” in Casey Rebecca Johnson (Ed.), *Voicing Dissent: The Ethics and Epistemology of Making Disagreement Public* (New York: Routledge).
- _____ (2019), *Just Words: On Speech and Hidden Harm* (Oxford: Oxford University Press).
- Mill, John Stuart (1991), “On liberty” in John Gray (Ed.), *On Liberty and Other Essays* (Oxford: Oxford University Press. [Originally published 1859])
- Schauer, Frederick (2015), *The Force of Law* (Cambridge MA: Harvard University Press).
- Sherwood, Merryn (2019), “Citizen journalists, sports fans, or advocates? The motivations of female independent sports media producers in Australia”, *Australian Journalism Review* 41 (2): 183-96.
- Stefancic, Jean, and Richard Delgado (1993), “A shifting balance: freedom of expression and hate-speech restriction”, *Iowa Law Review* 78 (3): 737-50.
- Tirrell, Lynne (2019), “Toxic misogyny and the limits of counterspeech”, *Fordham Law Review* 6 (87): 2433-452.
- Waldron, Jeremy (2012), *The Harm in Hate Speech* (Princeton: Princeton University Press).