Solving the Authority Problem: Why We Won't Debate You, Bro

Chris Cousens[1]

Abstract

Public arguments can be good or bad not only as a matter of logic, but also in the sense that speakers can *do* good or bad things with arguments. For example, hate speakers use public arguments to contribute to the subordination of their targets. But how can ordinary speakers acquire the authority to perform subordinating speech acts? This is the 'Authority Problem'. This paper defends a solution inspired by McGowan's (Australas J Philos 87:389–407, 2009) analysis of oppressive speech, including against concerns raised by McGowan (Just words: On speech and hidden harm, Oxford University Press, Oxford, 2019) herself. A *deflated* kind of authority can be gained from the hate speaker's standing in a norm-governed 'activity of oppression'. We should be wary about engaging with such speakers in public argument. Even if we counter their arguments, we may still elevate their standing within that activity and so enable them to perform more pernicious speech acts than was previously possible.

Keywords Speech acts · Authority · Oppression · Hate speech · No-platforming

1 Introduction

When philosophers talk about 'good' and 'bad' arguments, we tend to be looking for inconsistencies and contradictions. But arguments, especially public arguments, can be good or bad in other ways. An argument delivered by a prosecutor might lead to an innocent defendant being incarcerated. An argument from a politician might convince people to reject good laws. Public arguments can be bad even if the argument itself is 'good'. When thinking about the norms of public argument, we must include how such norms affect who argues, how they argue, and the conventional effects of their arguing.

This paper examines a specific way that public arguments can be bad. Public arguments are speech, and speech can oppress. Politicians vote 'Aye' for oppressive laws, employers instruct Human Resources departments to enact discriminatory hiring practices, and judges pronounce unjust sentences for minority defendants. Such speech straightforwardly contributes to oppression—powerful people change the rules to the disadvantage of a marginalised group. These speakers have the *authority* to enact such changes. Part of

being a politician is being able to enact new laws by voting. Part of being a judge is being able to pronounce sentences. Speakers with this 'positional' authority can perform special speech acts that everyday speakers cannot.

But everyday speakers, without any special authority, can *also* perform oppressive speech acts. Hate speech is not only bad because it conveys noxious ideas and inflicts psychological harm, although of course it does both. It also *constitutes* an act of oppression, changing social rules and norms in oppressive ways (McGowan 2009; Popa-Wyatt and Wyatt 2018; Tirrell 2012; Waldron 2012). Oppression is the unjust, systematic, structural subordination of a social group (see Frye 2000; Haslanger 2004; Matsuda 1993; McGowan 2009). Changing the rules is the kind of systematic and structural harm that fits this description—whether perpetrated by a politician through voting or an everyday speaker through hate speech.

But this brings us back to *authority*. The judge clearly has it when passing sentence, but the everyday hate speaker seems to be lacking it. This is the 'Authority Problem' (Maitra 2012). How can ordinary hate speakers change the rules for their target? My answer is that the norms governing oppressive social structures give members of privileged groups the (non-positional) authority required to perform oppressive speech acts. The white hate-speaker gains the

✉ Chris Cousens
   chrisjcousens@gmail.com

1   The University of Stirling, Stirling FK94LA, Scotland, UK

authority to speak oppressively in virtue of their (privileged) relationship to the 'activity of racial oppression'.

Not all members of a privileged social group have *the same* relationship with that activity; some hate speakers oppress more, and in different ways, than others (Tirrell 2018, see also Barnes forthcoming). One way that speakers can acquire a special relationship with an activity of oppression is to engage in public argument. Inviting a hate speaker to deliver a public lecture *elevates* them within that activity. It makes them the kind of person who gets invited to deliver controversial lectures at prestigious organisations. This can empower them to perform more harmful speech acts than was previously possible. We should be wary about debating or platforming such speakers—not because their *arguments* are compelling, but because arguing with them elevates their standing within an activity of oppression, even if they 'lose' the argument.

§2 outlines the 'Authority Problem', and §3 offers my solution: a 'deflationary' account of speaker authority drawing on McGowan's (2009) notion of 'rule-governed activities of oppression'. §4 defends this view against two potential objections, including one drawn from McGowan (2019) herself. §5 uses this argument to justify no-platforming and refusing to debate hate speakers.

2 The Authority Problem

Hate speech is supposed to not only *cause* oppression, but also to *constitute* an act of oppression (Langton 2012; Maitra 2012; McGowan 2009); similar arguments have been made for pornography (Langton 1993; Langton and West 1999) and slurs (Cousens 2020; Popa-Wyatt and Wyatt 2018). Oppression is, roughly, an unjust set of rules (and rule-like norms) constraining a target social group. Frye (2000) uses the metaphor of a birdcage. Lots of rules and norms constrain what people may or may not do; if enough are woven sufficiently tightly around a certain group, members of that group are 'caged'.
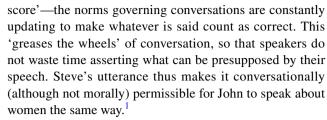
To show how seemingly ordinary utterances can constitute acts of oppression, McGowan (2009, p. 399) discusses the following example:

John: So, Steve, how'd it go last night?
Steve: I banged the bitch.
John: [smiling] She got a sistuh?

Steve's utterance changes the rules governing this conversation, making it appropriate to use denigrating language to describe women. It does this through 'accommodation' (Lewis 1979)—even if it is not normally appropriate to speak about women like this, Steve's utterance presupposes that it is, here and now, permissible to talk this way. The presupposition automatically enters the 'conversational score'—the norms governing conversations are constantly updating to make whatever is said count as correct. This 'greases the wheels' of conversation, so that speakers do not waste time asserting what can be presupposed by their speech. Steve's utterance thus makes it conversationally (although not morally) permissible for John to speak about women the same way.[1]

This change to the rules governing the conversation is a bit like adding one very tiny piece of wire to the birdcage of gender oppression (McGowan calls oppression a norm-governed 'activity'). As the total set of norms constituting the subordination of women now includes the rule that it is permissible in this conversation to denigrate women, it is *slightly* worse than it was before.

Some of the norms that make up an activity of oppression are strong and enduring. Men's sport is more 'exciting'. Women are more 'nurturing'. Boys will be boys. And so on. Such norms are developed incrementally over long periods of time. By contrast, norms like Steve's update to the conversational score are small and fleeting. It does not just change conversational rules, but also reinforces more enduring rules about the appropriate way to talk about women. As a result, oppression can be self-sustaining. Steve talks like this because of enduring (oppressive) social norms, and in doing so he reinforces them (Cousens 2020).

McGowan (2009) calls the kind of speech act Steve performs a 'covert exercitive'. Following Austin (1975), an exercitive is a speech act that changes permissibility conditions. Steve's utterance changes permissibility conditions by making it more acceptable to denigrate women. However, unlike overt exercitives such as 'You may leave now', Steve's exercitive move is *covert*; he sneaks the change to permissibility conditions into the conversational score by presupposing it.

This is where the Authority Problem appears. Granting and rescinding permission is typically tied to positional authority. Politicians can pass laws because a system of government gives them the authority to do so. Parents can set rules for their children because they have social (and usually legal) authority over them. Employers can determine the status of their employees because they have the authority (within that workplace) to do so.

But Steve has no formal positional authority. He is just some guy being awful. How can *he* change what is, and is not, permissible behaviour?[2]

---

[1] This is a compressed explanation of a complex phenomenon; see Langton (1993, 2018a; with West 1999), McGowan (2004, 2009), and Witek (2019) for more detail.

[2] The statement of the problem is kept brief; for more, see Maitra (2012) and Barnes (2016).

Authority is a felicity condition for many speech acts. One of Austin's (1975) examples is that only the designated namer can name a ship. Someone cannot leap in front of the namer, smash a champagne bottle over the bow, cry 'I name this ship the *Generalissimo Stalin*', and in doing so name the ship. As Kukla (2014) suggests, one cannot name babies by running through a maternity ward shouting out names. At face value, Steve's utterance might seem to go awry like this, failing to change permissibility because Steve lacks the authority to enact such changes.

McGowan argues that Steve's utterance *does* change permissibility conditions because some exercitive utterances—conversational exercitives—simply do not require the speaker to have authority. Anyone can do it. As McGowan (2019, p. 51) puts it, "Standard exercitives enact permissibility facts via an exercise of speaker authority. Conversational exercitives, by contrast, do so via the norm-governed nature of conversation." Conversational participants change permissibility conditions all the time. When John asks the question, he makes it appropriate (permissible) for Steve to answer. But not all permissibility-altering speech is oppressive.

What makes Steve's utterance *oppressive* is that it is not only a conversational move, but also a move in an 'activity of oppression'. Politicians voting in favour of oppressive legislation likewise make a (much larger) move in an activity of oppression. Their move (voting) requires special positional authority. Steve's move (covertly altering conversational permissibility) does not. But as both add to the set of norms that make up a very broad system of oppression, both constitute oppressive speech acts.

Cepollaro (2020) argues that this does not do justice to the 'special oppressive power' of hate speech. Hate speech is one of the worst things we do with our words; boiling this down to a commonplace conversational move might trivialise it. Furthermore (citing Bianchi 2014), some speakers can do *more oppressive* things with their words than others. But the move Steve makes in his utterance is supposed to be equally available to everyone. Cepollaro suggests that this model cannot explain why some speakers are systematically more authoritative than others.

An alternative solution to the Authority Problem is offered by Maitra (2012, p. 101). She considers the following example:

> An Arab woman is on a subway car crowded with people. An older white man walks up to her, and says "F***in' terrorist, go home. We don't need your kind here". He continues speaking in this manner to the woman, who doesn't respond… Many of the passengers turn to look at the speaker, but no one interferes.

Maitra argues that the speaker gains authority through 'licensing'; the speaker acts as if they have the authority required to subordinate the target (by *ranking* them as inferior, see Langton 1993). The silence of the other passengers is taken as tacit endorsement, granting the speaker the authority that they had assumed. In the same way, a group of friends might be trying to organise a hike. One of them takes charge, assigning tasks to the others. If everyone plays along, the speaker gains the authority required to assign tasks within the activity of hike-planning. The (tacit) acceptance of the audience *licenses* the authoritative speech of the speaker—although social hierarchies influence how easily someone can obtain licensed authority (Russell 2019).

Barnes (2016) and Bianchi (2019) argue that Maitra's solution does not connect correctly to the wider social context. An 'angry young man' could say 'We don't need your kind here' to a wealthy older person (Barnes 2016). Or a gay man might say 'We don't need your kind here!' to a heterosexual man (Bianchi 2019). These speakers attempt to subordinate a privileged group (wealthy old people or heterosexual men). If the audience does not interject, then the speaker is supposed to gain the authority required to subordinate through licensing and thus actually subordinates. But while we might offend them, we cannot *subordinate* wealthy older people or heterosexual men by shouting at them on the subway.

Barnes's alternative is to argue that everyday hate speakers leverage *informal situational positional authority*, which is positional authority *acquired* here-and-now. This is similar in spirit to Langton's (2018b) and Russell's (2019) solutions, which show how hate speakers might *gain* authority by presupposing it (accommodated by the audience). But *gaining* or *acquiring* authority leaves out something important: the sense in which privileged speakers *begin* conversations with the authority required for efficacious hate speech.

McGowan's and Maitra's solutions to the Authority Problem may be able to overcome the problems outlined above. However, my goal is to develop an alternative solution derived from another of McGowan's notions—the rules and norms governing an 'activity of oppression'. Rather than hate speakers *gaining* the authority to change permissibility (Barnes 2016; Langton 2018b; Maitra 2012; Russell 2019), or not needing it (McGowan 2019), I argue that they *already have* all the authority they need to oppress with their words.

3 Deflating Speaker Authority

To highlight why another solution to the Authority Problem might be needed, consider the following example:

> An Arab woman is the first person to arrive at a school board meeting. An older white man enters, sees her, and says "F***in' terrorist, go home. We don't need your kind here".
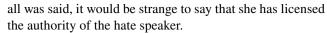
There is no audience to give the man's hate speech licensing as in Maitra's solution (Cepollaro 2020; Popa-Wyatt and Wyatt 2018). And while there is certainly covert norm-changing going on in the background (McGowan's solution) there is also an *overt*, foreground speech act to explain. The man attempts to instruct, or command, or order, the woman to leave. He *tells* her to go home. If he succeeds in *telling*, the woman has a reason to leave that she did not have before: she has been told to leave.

She will also have pragmatic reasons to leave. The situation is hostile, she is alone in a room with a hate speaker, and she may be justifiably concerned for her safety. On the other hand, leaving might feel like an uncomfortable concession to the hate speaker. Whether these considerations affect her decisions will depend on the participants and the context of the conversation. The illocutionary force of the hate speech underdetermines what happens next. But in addition to these practical concerns, the utterance reshapes the normative landscape such that there is at least some pressure on her to leave *in virtue of being told to go*.

This normative pressure is defeasible. The woman might *block* the attempted telling, in Langton's (2018a) sense: 'You have no right to tell me to leave!'. She might pointedly ignore it, denying it the public ratification that Kukla (2014) argues is required for the full performance of a speech act. She might try to *change* the speech act performed by saying 'Thanks for the suggestion, but I'd prefer to stay' (Caponetto and Cepollaro 2022; Cousens m.s.).[3] But before any response, the hate speech gives the woman a speech-act-related reason to leave. She might explain a decision to do so in those terms: 'He told me to leave, and rather than make a scene I decided to go.'

If this is correct and the man has succeeded in *telling*, the Authority Problem applies. What gives him the authority to tell the woman to leave? He does not have positional or formal authority, and the woman has every right to be there. And the *telling* is not a conversational exercitive, presupposed and accommodated. It is right there in the locution: 'go home'. He cannot gain licensed authority from silent bystanders; no one else is present. The woman might provide something like licensed authority herself if she 'plays along' and leaves or otherwise tacitly accepts his presupposition of authority (invoking Langton 2018b and Russell 2019). But if she stays, perhaps glaring defiantly at the speaker, perhaps deploying counter-speech, perhaps behaving as if nothing at

all was said, it would be strange to say that she has licensed the authority of the hate speaker.

I suggest that the hate speaker has the 'authority' to tell the Arab woman to leave in virtue of his standing within the activity of racial oppression. He occupies at least two privileged roles in separate oppressive domains: a man in a patriarchy, and a white person in a racist society. Here I focus on the racial domain, although it cannot be so easily abstracted in real-world interactions. Whitten (2019, p. 566) argues something similar when analysing such an example: a hate speaker attempts to "assert his authority in his role as a white, British person"—this also derives the authority to oppress from group membership.

Hate speech is supposed to *oppress* its targets by changing local permissibility conditions in an oppressive way. Changing permissibility is the domain of *exercitive* language—like judges passing sentences and parents forbidding things for their children. In these paradigmatic cases, speakers have 'positional' authority underwriting their exercitive speech, bestowed on them by the legal system and legislation respectively. But hate speech is delivered by many people who do *not* have positional (or institutional) authority. How can *they* change what is permissible for their targets? McGowan (2019) says that conversational (covert) exercitives do not require authority—they are 'parallel' speech acts which have different felicity conditions to 'Austinian' speech acts (like sentencing). Barnes (2016), Langton (2018b), Maitra (2012), and Russell (2019) suggest that speakers might acquire something akin to positional authority from their audience. In this section, I argue for an alternative solution. Subordinating speech acts *do* require authority—not *positional* authority but a more mundane *speaker authority*—being the 'right' person to perform a certain (oppressive) speech act.

There are many ways that speakers can gain the authority required to perform a speech act. The judge gains the authority to pronounce sentences from their position within the legal system. The ship-namer has the authority to name their ship in virtue of owning it. The conventions governing sentencing and ship-naming identify the *right* person to perform that speech act. There is no Authority Problem here.

This extends to pretty much any speech act. Speech acts are governed by "conventional procedures", and this includes "the particular persons" who might appropriately perform them (Austin's 'A1' and 'A2' rules in his 1975, p. 14–15). While some speech acts may be performed by specific individuals (only the ship-owner may name their ship), others are much wider in scope. The 'right' person to make a promise is the person who is committing to future action. The 'right' person to place a bet is the person who is putting forward the money. Speakers gain the authority to perform these speech acts in virtue of their standing within a rule-governed activity (naming, promising, or betting). This

---

[3] Cousens, C. (m.s.) 'Exercising Illocutionary Power: Or, how to do things with other people's words', in Popa-Wyatt, M (ed) Harmful Speech and Contestation, Palgrave MacMillan (forthcoming 2023). These strategies attempt to subvert the illocutionary force of the hate speech, but not (directly) its pernicious *content*—so are not traditional forms of counter-speech.

'speaker authority' is, I suggest, simply the fact that someone has been picked out by the conventions of an activity as the 'right' person to perform a speech act.

There are contrasts between 'speaker' authority and the 'positional' authority normally associated with exercitive speech acts. Positional authority is usually institutional—a judge gains their positional authority from the institution of the legal system. The normative changes they make tend to be binary—if the judge prohibits something, then it simply becomes impermissible. And their authoritative utterances are binding regardless of whether a defendant recognises their authority. On the other hand, speaker authority need not be backed up by formal institutions—there is no regulation or legislation governing who can make promises. These normative changes can be 'gradable'—promises can be stronger or weaker. And the success of a promise is more responsive to the audience ('There's no need to *promise* me!').

The features of informality, gradeability, and audience-responsiveness are part of a general contrast between speaker and positional authority, not a set of necessary conditions. In non-ideal contexts, positional authority will probably share some of these features. But even though 'speaker authority' might at times seem tied to informal social *positions*, I hope to have sketched out how it can work differently to positional authority—but it can still change permissibility. Promising makes actions that would break the promise impermissible.

For many speech acts, it is straightforward who the 'right' person is and how they are related to the wider activity (the judge to the legal system, the promiser to promising, etc.). It is less clear how someone can be picked out by the conventions governing an activity of oppression as the 'right' person to perform a subordinating speech act. I will try to sketch this out in a little more detail.

Oppression, like many human activities, is governed by lots of rule-like norms and conventions (McGowan 2009). These include, but are not limited to, beliefs (such as the belief that Arab people are terrorists), norms (such as the norm that one should exclude Arab people from public spaces), and practices (such as harassing Arab people). Hom (2008) suggests that racist epithets are underpinned by similar 'institutions of discrimination'.

The white man's utterance in the school board example is oppressive. Telling a member of a subordinated group to leave a public space is not merely a conversational 'telling', but also a move in an activity of racial oppression. By racializing his 'telling', the speaker is not only changing permissibility facts for the woman; he is trying to change them for other members of the targeted group as well. If the woman should 'go home' because of her race, then so should others.

As the hate speech contributes to the wider activity of racial oppression, it constitutes an oppressive speech act.[4]

Some people can make this move successfully; others cannot. In the most straightforward instances, white people are picked out by the rules governing racial oppression as the 'right' speakers to perform oppressive speech acts. To successfully perform the oppressive action of *telling an Arab to 'Go home!'*, the speaker should be white. Something would seem to go awry if the speaker was Arab themselves, as they would not be the 'right' person to perform that oppressive speech act. In-group members can certainly contribute to their own subordination (e.g., women who campaign against abortion rights). But this is a different *kind* of move in the activity of oppression. When the move is performed with speech, we might say it is a (subtly) different speech act. If a non-white, non-Arab speaker uttered the hate speech, it might again be best explained as a (subtly) different oppressive move to that performed by the white hate-speaker.

When the hate speaker speaks, they have the *speaker* authority required to felicitously *tell* the target to 'go home'. The speech act of *telling an Arab to 'Go home!'* is part of a wider activity of racial oppression (one of many possible utterances that police the use of public space by marginalised groups). The conventions governing that speech act include identifying the people who can perform it (like the conventions of marriage determine who can perform the act of marrying). In this case, the speakers able to paradigmatically perform that speech act are white people. The white speaker does not have positional authority (like a judge), but are nonetheless, according to the norms governing the activity of racial oppression and its related speech acts, the 'right' person to reshape norms governing a public space in racially oppressive ways.

This would solve the Authority Problem. Speakers acquire the authority needed to perform speech acts in virtue of their relationship to a relevant activity. Celebrants gain the authority required to perform marriages in virtue of their relationship to the activity of marriage. And hate speakers gain the authority required to perform subordinating speech acts in virtue of their relationship to an activity of oppression. This is a 'deflated' sense of authority—the speaker does not have positional authority *over* their target, but still has the lesser authority required to perform an oppressive speech act.

One might think that the attempted speech act simply fails. The woman is not, in fact, *told* to go home because the man has no authority to tell her to go. He is trying to

---

4 Russell (2019) describes something similar—subordinating speech acts—although seems to take these to be closer to Austinian exercitives than McGowan (or I).

change permissibility facts about who gets to access this public space, but without authority, his attempt fails.

This would misunderstand how fine-grained permissibility can be. In straightforward examples, permissibility is binary. Something is, or is not, allowed. A parent might say 'If you don't eat your meat, you can't have any pudding!' to set a clear permissibility condition for their child (see also Lewis's (1979) example of the master and slave). But permissibility is gradable. Stealing and murder are both impermissible, but the edict against (and sanctions attached to) the latter are much stronger. I think it makes sense to say that murder is *more* impermissible than stealing. Furthermore, permissibility can be vague. A parent says to their child 'Don't drink too much!'; how much is too much? This will vary according to the context. If said to two children, it might permit the older child a fourth drink but require the younger to stop at two.

Permissibility facts about something as complex as access to public space will not be as clear-cut as a judge's pronouncement of sentence. The hate speaker does not make it *impermissible* for the woman to remain; he might still make it *less permissible*, or less appropriate. This is still a change to permissibility conditions. Before the hate speech, the woman may obviously remain in the room. After the hate speech, it is *less* appropriate for her to do so than it was before (this is descriptive—the *moral* inappropriateness is of the hate speech itself).

It is not only permissibility conditions that are shifted by hate speech. Here, I collapse a broad set of norms—governing politeness, social coordination, 'face-threatening' behaviour, racism, conversational kinematics, and more—into a single question of permissibility, as this framing addresses the Authority Problem. It is a shame to gloss over the many ways that we might describe the normative changes enacted by something as politically and psychologically charged as hate speech, but there is only so much that one paper can explore.

The idea I am trying to tease out is that hate speech does not need to change permissibility conditions from 'stay' to 'go' to be oppressive, but rather from 'clearly stay' to 'slightly-less-clearly stay'. The speaker at the school board meeting updates permissibility conditions not by *reversing* permissibility but by changing the *strength* of permissibility.

So, *positional* authority is not required for everyday hate speakers to subordinate with their words. They do not need the same kind of authority as the parent or judge because they are not enacting the same *kind* of norm-change. Rather than issuing a binding (and binary) edict, they leverage their privileged group membership in an oppressive social hierarchy to shift the strength of permissibility conditions. The white man at the meeting does not make staying *im*permissible for the woman, he just makes it *less* permissible. But this is still a change to the overall set of rules and norms that

makes up the activity of racial oppression, and so it constitutes an (albeit small) oppressive speech act.

As noted above, this is something of a sketch. I finish this section with some additional detail.

Firstly, this suggests a different kind of authority from that of the judge or legislator. The judge has clear positional authority and can perform legally binding speech acts. The ordinary speaker has none of this. Maybe 'authority' is the wrong word to use, and something more general like 'power' would better suit (Simpson 2021)? However, Langton (2018b) suggests that there are at least two ways that we use the term 'authority'. We can have authority over someone—'practical authority', and we can also be 'an authority' about something—'epistemic authority'.[5] Roughly, someone has epistemic authority if their saying something means I should believe it; they have practical authority if their telling me to do something means I should do it. While I do not think that the authority required to perform oppressive speech acts *is* epistemic authority, it influences reasoning in a similar way.

Simpson (2021) suggests that for a speaker to have authority, their speech should be authoritative—lawlike in a way that small changes to the strength of permissibility conditions are not (see also Bauer 2015; Raz 1986). When A has authority over B, A's authoritative speech should *displace* B's own reasons for acting. The parent who says to the child 'You cannot touch that!' displaces the child's own reasons for touching that with a new reason to not touch it. On the other hand, subway hate speech telling someone to *go* does not displace their reasons to *stay*.

However, we often describe people as 'authoritative' in a more general sense, such as an expert being 'an authority' on a topic (epistemic authority). We do not mean that they have authority *over us* and that their judgements will 'displace' our own reasons—it is more a recognition of their special standing in a norm-governed activity. For example, a literature professor who is an expert on Lord of the Rings is an 'authority' on that topic. If they were to say that Bilbo, not Frodo, destroys the Ring of Power, their utterance would not *displace* my reason for believing otherwise (that I have read the book myself). I would assume that they had misspoken. But it still weighs into my reasoning, perhaps leading me to double-check the book or ask them for clarification.[6]

This parallels the role played by speaker authority in oppressive speech acts. The hate speaker does not displace their target's reasons for acting, but rather shifts normative expectations a bit (or a lot). This weighs into reasoning about what to do (stay or go), but not in the same way that

---

[5] These can be intertwined, see also Lewiński (2022) and Raz (2009).

[6] Arguably it was Gollum, not Frodo, who destroyed the ring; action theorists may need to adjudicate.

an instruction from a parent would for their child. So, not all authority connected to social hierarchies needs to be reason-displacing in Simpson's strict sense of the term.

Secondly, this ties the authority required to perform subordinating speech acts to 'broader social norms'—avoiding one of Barnes's (2016) and Bianchi's (2019) objections to Maitra's (2012) solution. The authority I describe here is gained from an activity of oppression, rather than conversational mechanisms. Deflating 'speaker authority' suggests that it is nothing special; many people can perform subordinating speech acts. Speakers without positional authority can still nudge norms in more (or less) oppressive directions, subtly changing the overall set of permissibility conditions that are part of an activity of oppression.

This proposed solution also generates the right analysis of Barnes's (2016) and Bianchi's (2019) examples of attempted subordination of privileged groups. There *is* no norm-governed 'activity of oppression' of wealthy older people, or of heterosexuals. Even if the speaker succeeds in ranking the target group as inferior in the conversational score, their utterance cannot be a move in an activity of oppression that does not exist (Bianchi 2019 suggests something similar, p. 413).

This section advocates re-thinking the 'authority' in the Authority Problem. Even if traditional exercitives require *positional* authority, deflated *speaker* authority—being picked out as the 'right' person to perform a norm-altering speech act—can allow ordinary hate speakers to subordinate with their speech. There is not space here to tease out all the implications of this deflationary project, but I hope to have shown how this can solve the Authority Problem and explain how speakers oppress with their words.

4 Objections

In this section, I consider two potential objections to this McGowan-inspired solution, the first drawing on McGowan's own work.

In *Just Words*, McGowan argues that attributing authority too widely might not account for the 'special oppressive power' of hate speech—"To say that all speech acts are authoritative speech acts is to lose sight of the important role that power plays in only some types of speech act" (McGowan 2019, p. 81). And my proposed solution does suggest that all speech acts require 'deflated' speaker authority (being the right person to perform it), although I say only that McGowan *might* object to my view in this way. My description of speaker authority might be quite close to McGowan's notion of the 'standing' required for conversational contribution (although I think 'authority' better captures the relationship between hate speaker and social norms). Cepollaro (2020) offers a similar objection to

McGowan's argument that hate speech enacts permissibility conditions without requiring authority. Hate speech is explosively harmful and deflating the importance of the authority required to perform subordinating speech acts might not do justice to that harmfulness.

My pre-emptive response to this kind of objection comes in two parts. Firstly, a deflationary solution to the Authority Problem leaves all the psychological harms of hate speech intact. The speaker on the subway inflicts the same emotional violence whether their authority is gained from their standing in the activity of oppression, the licensing of the audience, or without authority at all. The most immediate harms of hate speech are just as explosive, on my account, as any other.

Secondly, hate speech in a *legal* sense (McGowan's focus) needs to be explosively vile to overcome the general 'right' to free speech and be justifiably restricted (see e.g., Feinberg 1985). However, speech which falls short of a legal standard of hate speech can still offend, intimidate, and humiliate. It can also alter the norms governing a speech situation to the detriment of the target group—in other words, speech can be oppressive even if it would not fit a *legal* definition of hate speech. Even if all hate speech is explosive and subordinates, not all subordinating speech is (legally classified as) hate speech; explanations of subordinating speech do not need to treat all such speech as explosive. If my solution to the Authority Problem concludes that some inexplosive utterances are oppressive, that should count in its favour.

For example, the speaker on Maitra's subway car might say pointedly 'I think that Arabs are terrorists and that they should not be allowed on the subway with us.' This is not the formulation of racist speech that typically mirrors 'fighting words' (Matsuda 1993). As it is phrased as a statement of opinion, it might count as protected political speech. Nonetheless, it would still change the normative landscape in a similar way to the instruction 'go home'. It *suggests*, or *implies*, that the woman should leave, even if it does not *tell* her to go.[7] This still puts pressure on the woman to leave. Even though it is less explosive, and perhaps not regulable as hate speech, it functions similarly to subordinate and oppress—albeit plausibly as a *less* oppressive move. There are many ways that privileged speakers can shift the norms constituting an activity of oppression to make things worse for those at the other end of the hierarchy. If our analysis is restricted to only the very worst instances of hate speech and the authority required to perform these, we might overlook

---

[7] This is not to say that a suggestion is simply a weak command—different speech acts can impose different *kinds* of obligation (Lance and Kukla 2013). For my purposes here, glossing this as 'stronger' and 'weaker' obligations works well enough.
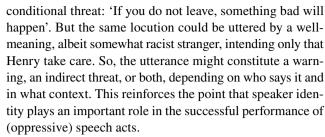
the nuanced ways that oppression is propagated at the conversational micro-level (e.g., McClure 2020).

A second possible objection comes from Bianchi (2014), echoed by Cepollaro (2020). If the activity of racial oppression picks out white people as the 'right' speakers to perform oppressive speech acts, then it would seem to suggest that all white people have the *same* power to alter permissibility facts. But some speakers are systematically more authoritative than others. Langton (1993, p. 299) says something similar: "If you are powerful, there are more things you can do with your words". A radio host with a large following might be able to alter permissibility facts much more than an unpopular member of a small white supremacist group. This is posed as a problem for McGowan, who argues that no special authority is required to use oppressive covert exercitives, but it would also be a problem for me. If all white speakers have the authority to speak oppressively in virtue of being white in a racist society, why can some white people speak so much more oppressively than others?

There are two parts to my answer, although the precise mechanics are a little tentative (to keep options open). Firstly, there are lots of different kinds of oppressive speech acts ordinary speakers can perform. They might *warn* members of the target group to avoid a public space ('Your kind should be careful hanging around here.'). They might *instruct* them to leave ('Fuck off, terrorist!'). They might *undermine* their positional authority ('Real cops don't wear turbans!). Different kinds of authority are required for traditional speech acts—a parent instructing a child needs a different kind of authority from a celebrant officiating a marriage or a gambler placing a bet. Likewise, different oppressive speech acts might pick out subsets of a privileged group as the 'right' people to perform them.

For example, imagine a white speaker trying to issue an oppressive warning to Mark Henry, professional wrestler, and 'World's Strongest Man', by saying 'Your kind should be careful hanging around here.' If the speaker is a child, or alone, Henry may well feel no normative pull to leave *at all* (although this would be a powerful demonstration of the underlying oppressive character of this social context). The attempted oppressive warning fails. But if the speaker is adult, perhaps part of a group, or perhaps armed, Henry might feel that leaving is now more appropriate. *Who* attempts to perform the speech act matters, in more ways than just being white. If this is right, the speech act of an *oppressive warning* is different to a traditional Austinian warning which anyone can perform, as the speaker needs to not only be white (in this instance) but to also have some additional kind of power, threat, or special standing within the activity of oppression.

This might be construed as a kind of veiled threat. Threats are conditional (Schiller 2021); maybe the speaker uses the surface grammar of a warning to issue an indirect conditional threat: 'If you do not leave, something bad will happen'. But the same locution could be uttered by a well-meaning, albeit somewhat racist stranger, intending only that Henry take care. So, the utterance might constitute a warning, an indirect threat, or both, depending on who says it and in what context. This reinforces the point that speaker identity plays an important role in the successful performance of (oppressive) speech acts.

This is the second part of the answer. There are many relations speakers can have to an activity of oppression, just as there are lots of different kinds of white person. Some occupy positions of formal authority (politicians or judges) while others do not. But even amongst those without formal authority, different members of an oppressor group can perform different oppressive speech acts.

As noted above, subordinating speech is not a single speech act that can be performed by any white speaker in virtue of being white within an activity of racial oppression. This activity is complex and includes many oppressive speech acts, each with its own felicity conditions identifying the 'right' people to perform it. The norms governing these speech acts are the kind of *enduring* norm that was mentioned in §2, and so exist prior to the performance of a specific token of oppressive speech. Ordinary speakers can reinforce (or undermine) such norms through their conversational contributions (as Steve does in McGowan's 2009 example), but changing such norms happens incrementally. Langton (2018b, p128) similarly notes that hate speech can "normalise itself".

Some oppressive speech acts may be equally available to all members of a privileged group (such as slurring). But others might only be able to be performed by some (such as oppressive warning). Some speakers can, in virtue of their role in an activity of oppression, change permissibility more easily than others (see Russell 2019 for an alternative explanation of the same phenomenon). Whitten (2019) likewise suggests that 'community leaders' can, even without *positional* authority, shift norms more easily than the 'members' of the community.

Speaker identity can thus affect the success (or failure) of an attempted oppressive speech act (see also Tirrell 2018). But it can also mean that speakers perform different speech acts with the same words. Imagine three white supremacists speaking at a rally. The first, Tucker, is a popular TV host. The second, Richard, is a notorious white supremacist and former leader of a hate group. The third, Patricia, is an ordinary and unremarkable bigot. [8].

If each uttered the phrase 'See how they disrespect us? It's time to take revenge for 9/11!', a different speech act

---

[8] While references to real people are intentional, this example is fictional.

might be performed. Patricia's utterance might function as a *suggestion*, something that the gathered white supremacists think is a good idea, and some decide to act upon. Richard's utterance works more strongly, perhaps as an *instruction*. The audience might not decide to follow that instruction, but they feel the pull of it more fiercely than when Patricia speaks. Richard's utterance, despite using the same words, could impose a heavier normative burden on the audience because he has an elevated standing in the activity of racial oppression (despite not having *positional* authority). Tucker's utterance could be stronger still, taken to be a *command* or *order* by an audience accustomed to believing and doing whatever their favourite TV host says. Even if Patricia was *trying* to command the audience, because of her lower standing she is not the right person to issue a command, and so her attempt may fail.

It might be that Tucker imposes a *stronger* normative burden on the audience (as Caponetto 2020 characterises the amendment of illocutionary force). Or maybe each imposes a different *kind* of obligation (Lance and Kukla 2013). This raises questions about the precise mechanics of speech act theory. If illocutionary force is determined by speaker intention, then we could simply ask the speaker which speech act they were trying to perform, and the audience if they recognised it (something like this view is often attributed to Austin 1975, see McDonald 2022). But even if the speakers are intending to perform the same speech act, their utterances might have different illocutionary forces. One suggests, one instructs, another commands. This might make it seem that the audience's *uptake* of a speech act determines its character, even if it leads to the performance of a speech act that the speaker did not intend (Kukla 2014).

But, as McDonald (2022) argues, this seems to diminish the role of speakers too far. McDonald suggests an in-between position, where speaker and audience *collaborate* to determine the illocutionary force of an utterance through their subsequent behaviour. I think this is on the right track. But imagine a speaker in a video-call who says 'I promise to φ' a moment before the call cuts out. A 'collaboration' theory of uptake would struggle to explain the illocutionary force of the utterance without subsequent collaborative behaviour. I think the speaker *does* promise, perhaps non-ideally (McDonald only argues that collaboration is *sometimes* the best explanation of uptake). I suspect that a solution might be to take intention and uptake 'out of the heads' of speaker and audience, although must leave a full account of this for future work. For now, all that I require is that uptake is *more* than merely recognising the speaker's intention, but *less* than fully determining an utterance's illocutionary force.

So, three different white speakers utter the same locution in the same location, and each does something different with their words despite none having positional authority in this context. I am suggesting that because they stand in different relationships to the activity of oppression, each enacts a different change to the permissibility conditions governing the speech situation, and so each performs a different oppressive speech act.

Barnes (2016) describes something a bit like this notion of 'relationship to an activity of oppression'. He suggests that speakers 'create' or 'move into' *social roles* that give the occupier of that role informal, localised, positional authority. This could be an alternative way to explain the different oppressive illocutionary potential of white speakers. White speakers can adopt different *roles* within an activity of racial oppression, and in doing so gain the authority, here-and-now, to perform oppressive speech acts.

This could be made compatible with my account, but I think there are at least two reasons to prefer mine. The first is that a deflationary account of authority is an advantage. Even on Barnes's account, the kind of authority being acquired is 'authority over someone'. But this is a strange way to describe, for example, a poor white speaker aiming hate speech at a Black president. Instead, I think we should describe that speaker as having the *speaker* authority required to shift permissibility conditions in a way that subordinates Black people generally, rather than the president individually.

Secondly, on my account speakers do not need to 'move into' or 'create' a role that lends them additional authority. They have a lesser kind of authority, but they bring it with them to each conversation. There is no need to acquire it, and the fact that they *already* have it is a feature of an oppressive status quo. As Barnes says, for Maitra's subway hate speaker, the public space is more *his* than *hers*—even before he says anything.

5 Why We Won't Debate You, Bro

Different speakers stand in different relations to an activity of racial oppression. Some can change permissibility conditions more, and in more pernicious ways, than others. In the previous example, it was stipulated that Tucker, Richard, and Patricia each have a different standing in the activity of racial oppression, and this enabled some to do worse things with their words. But how does someone gain an elevated standing within such an activity?

One way that hate speakers can elevate their standing within an activity of oppression is as an unintended consequence of public argument. When confronted by hate speakers, we might want to debate them to show their audience how wrong they are. We might invite the hate speaker to deliver a public lecture, to expose their bad ideas. However, engaging with oppressive speech can *backfire*. Giving racist hate speakers a public platform changes their role in the
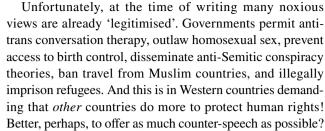
activity of oppression: they are now the kind of person who gets invited to talk about white supremacy.

This is a constraint on the efficacy of counter-speech—Lepoutre (2019) surveys others and concludes that proactive positive counter-speech is the most promising. Rather than respond *after* the hate speech, powerful actors (like governments, or universities) might try to say nice things about groups likely to be targeted with hate speech. But once the speaker has already been invited to talk, positive proactive counter-speech is likely to seem hypocritical—the institution has already invited, platformed, and profited from the hate speaker.

This section looks at two related moves in public argument: no-platforming and refusal to debate. The former has received a lot of attention—speakers are invited to give a talk at an institution (typically a university), there is a public protest, and the invitation is then (sometimes) rescinded. The speaker is denied a platform from which to speak, hence 'no-platforming'. This is different to 'cancelling' a speaker. A single event is cancelled, whereas 'cancelling' a controversial person is an ongoing attempt by a large group of people to reduce their influence in public life (so a token no-platforming decision might be *part* of an attempt to 'cancel' someone).

Refusal to engage is also an important move in public argumentation. A Palestinian activist might refuse to participate in a panel discussing the occupation of Palestine if the panel is not allowed to discuss the Boycott, Divestment, and Sanctions movement (BDS). A trans woman might refuse to write an opinion article for a newspaper replying to a transphobic article published the previous day. An academic might refuse to respond to an online troll asking for them to 'debate' him.

Why retract a platform from a potential hate speaker, or refuse to debate them? The justification for this is usually cached out in terms of either the psychological harms predicted to follow that person's speech or worries about 'legitimising' hateful ideas (Levy 2019; Peters and Nottelmann 2021; Simpson and Srinivasan 2018). Levy (2019, p. 487) sums the second thought up nicely: "Provision of a platform provides higher-order evidence that the view being argued for is worth taking seriously". Allowing a hate speaker to speak from behind a lectern emblazoned with the crest of a prestigious university suggests that they are worth listening to. Engaging with them in public debate suggests that the rights of their target are, in fact, up for debate. So, one reason often given for no-platforming and refusing debate is to avoid legitimising ideas that are *not* worth listening to.[9]

Unfortunately, at the time of writing many noxious views are already 'legitimised'. Governments permit anti-trans conversation therapy, outlaw homosexual sex, prevent access to birth control, disseminate anti-Semitic conspiracy theories, ban travel from Muslim countries, and illegally imprison refugees. And this is in Western countries demanding that *other* countries do more to protect human rights! Better, perhaps, to offer as much counter-speech as possible?

I set aside the 'epistemic' considerations for no-platforming, which have been discussed elsewhere (see Peters and Nottelmann 2021). There is another way that providing a platform to, or debating, hate speakers might be bad. Inviting a hate speaker to give a lecture, or engaging with them in public debate, changes their relationship to the broader activity of oppression in which they participate.

For example, a hate speaker who appears regularly on television shows and is invited to speak at white supremacist rallies will have an elevated standing within the white supremacist movement. They will be able to perform different types of speech acts to a bigot who posts on a poorly followed Twitter account. The regularly platformed speaker will be able to *call* to action, *incite* civil disobedience, and *instruct* others to further their agenda. An unknown bigot may be able to perform some of these speech acts. They could successfully call on others to perform a certain action, or (as in a previous example) tell an Arab woman to leave a school board meeting. But they would be unlikely to succeed in *commanding* others to act, or to *incite* civil disobedience. Even if this is their intention, their lower standing within the activity of oppression makes it harder for them to secure uptake. It would be like a child who tries to *command* their parents but fails to impose the normative burden of a command even if their attempt is recognised as such.

This is not a difference in perlocutionary effects, with one speaker enraging a larger crowd than the other. Tucker's utterance *constitutes* a different speech act to Richard's and Patricia's. If it secures uptake as a command, then the normative pressure on audience members to act in accordance with it is much stronger than if it secures uptake only as a suggestion. When Tucker speaks, the normative landscape changes in different ways to when Richard, or Patricia, speaks.

This will presumably also lead to different perlocutionary effects: a greater number of people trying to 'take revenge' because of the stronger normative burden of the command. Not all those who hear the command will follow it, but (presumably) more will than when the utterance is merely a suggestion. These perlocutionary differences are explained by illocutionary differences, and the illocutionary differences are partially determined by the standings of the speakers within the relevant activity of oppression. Only some are the 'right' people to perform more oppressive speech acts (like a command to commit hate crimes).

---

[9] Some of what follows extends beyond paradigmatic hate speech. Anti-vaccination arguments are often moves in the activity of ableism; holocaust denial is a move in the activity of anti-Semitism. No-platforming climate-change sceptics, on the other hand, might require *epistemic* justification.

Giving hate speakers a platform, or publicly engaging with their speech, offers them illocutionary 'power'—as this elevates their standing within an activity of oppression, their speech acts can place stronger normative burdens on their audience. Barnes (forthcoming) explains how a supportive audience can 'elevate' ordinary speakers—I am arguing that *un*supportive audiences can inadvertently do the same.

Levy (2019, p. 496) describes something similar: "provision of a platform provides some degree of certification of representativeness, as well as of credibility". 'Representativeness' is the implication that the speaker has been given this speaking opportunity because they are *not* an idiosyncratic crank. When someone with expertise (perhaps an academic) engages a hate speaker in public debate, they are (perhaps unintentionally) signalling that the person they are debating represents a widely shared viewpoint, one that *deserves* or *demands* academic engagement.

This is most pronounced when a debate is hosted by, say, a reputable university. But it extends to less prestigious venues such as Twitter. When an academic responds to an online troll, it shines a spotlight on that troll as representative of their point of view. Algorithms and internet search functions reward 'engagement'—the more something is clicked on, or the more likes it gets, the more often it shows up for other users. So, responding to the troll makes their hate speech more likely to be seen by more people. As their notoriety grows, so does their standing within their activity of oppression. And online debates are permanent. Conversational speech disappears, and changes in the conversational score are supposed to end with the conversation (Popa-Wyatt and Wyatt 2018). By contrast, a screenshot of a tweet can outlive the tweet itself, continuing to elevate the standing of the hate tweeter even after the original tweet is deleted (Waldron 2012 and Lai 2020 describe other kinds of permanent hate speech).

Working out whether to no-platform or refuse to debate a hate speaker is difficult. Peters and Nottelmann (2021) describe this as a 'dilemma'. One must weigh up the psychological harm that the speaker might inflict, the importance of 'free speech', the anticipated backlash from angry supporters, and the potential for effective counter-speech. No-platforming might even elevate the speaker within an activity of oppression *more* than permitting a forgettable public address, turning them into a martyr for their hateful cause. If debating or platforming a hate speaker enables them to change the normative landscape in ways that they could not before, then an attempt to defeat bad speech with good speech might backfire. But so can no-platforming them! Whether a no-platforming decision will do more harm than good is a tricky empirical question and will vary across different contexts.

6 Conclusion

The 'Authority Problem' can be solved by deflating the conception of 'authority' required to perform permissibility-altering speech acts. These do not always require *positional* authority, but rather a kind of informal, non-positional *speaker* authority. The conventions governing oppressive speech acts identify who can perform them, just as the conventions governing ship-naming identifies who can (felicitously) name a ship. But this does not mean that all members of an oppressor group can perform the *same* oppressive speech acts, even if their ability to perniciously affect permissibility conditions stems from the same 'activity of oppression'. Some hate speakers can perform worse speech acts than others.

This means we should be careful about who we platform or debate. Even though hate speakers make bad arguments, platforming them gives them a special representative role within an activity of oppression, lending their hate speech greater illocutionary power. Decisions about no-platforming and refusing to debate provocateurs are already difficult, given the potential for backlash and escalation of fractious situations. Nonetheless, if we are to make *good* decisions about how we conduct public argument, we need to consider the potential elevation of individual hate speakers within activities of oppression, otherwise we might inadvertently enable worse oppressive speech acts than counter-speech can combat.

Data Availability No new data or data sets were generated or analysed for this paper.

Declarations

Conflict of interest There are no conflicts of interest or funding related to this work.

References

Austin JL (1975) How to do things with words, 2nd edn. Oxford University Press, Oxford

Barnes MR (2016) Speaking with (subordinating) authority. Soc Theory Pract 42(2):240–257

Barnes MR (forthcoming) Presupposition and Propaganda: A Socially Extended Analysis. In: Caponetto L, Labinaz P (eds) Sbisà on Speech as Action. Palgrave Macmillan, London

Bauer N (2015) How to do things with pornography. Harvard University Press

Bianchi C (2014) The speech acts account of derogatory epithets: some critical notes. In: Dutant J, Fassio D, Meylan A (eds) Liber Amicorum Pascal Engel. University of Geneva, Geneva, pp 465–480

Bianchi C (2019) Asymmetrical conversations: acts of subordination and the authority problem. Grazer Philosophische Studien 96(3):401–418

Caponetto L (2020) Undoing things with words. Synthese 197(6):2399–2414

Caponetto L, Cepollaro B (2022) Bending as counterspeech. Ethical Theory Moral Pract. https://doi.org/10.1007/s10677-022-10334-4

Cepollaro B (2020) Slurs and thick terms: when language encodes values. Lexington Books, Pennsylvania

Cousens C (2020) Are ableist insults secretly slurs? Lang Sci 77:101252

Feinberg J (1985) Offence to others: the moral limits of the criminal law. Oxford University Press, Oxford

Frye M (2000) Oppression. In: Minas A (ed) Gender basics: feminist perspectives on women and men. Wadsworth, pp 10–16

Haslanger S (2004) Oppressions: Racial And Other. In: Levine M, Pataki T (eds) Racism in Mind. Cornell University Press, pp 97–123

Hom C (2008) The semantics of racial epithets. J Philos 105(8):416–440

Kukla R (2014) Performative force, convention, and discursive injustice. Hypatia 29(2):440–457

Lai T (2020) Political vandalism as counter-speech: a defense of defacing and destroying tainted monuments. Eur J Philos 28(3):602–616

Lance M, Kukla R (2013) Leave the gun; take the cannoli! The pragmatic topography of second-person calls. Ethics 123(3):456–478

Langton R (1993) Speech acts and unspeakable acts. Philos Public Aff 22(4):293–330

Langton R (2012) Beyond belief: Pragmatics in hate speech and pornography. In: Maitra I, McGowan MK (eds) Speech and harm: controversies over free speech. Oxford University Press, Oxford, pp 72–93

Langton R (2018a) Blocking as counter-speech. In: Harris D, Fogal D, Moss M (eds) New work on speech acts. Oxford University Press, pp 144–156

Langton R (2018b) The authority of hate speech. In: Gardner J, Green L, Leiter B (eds) Oxford Studies in philosophy of law, vol 3. Oxford University Press, pp 123–152

Langton R, West C (1999) Scorekeeping in a pornographic language game. Australas J Philos 77(3):303–319

Lepoutre M (2019) Can "more speech" counter ignorant speech? J Ethics Soc Philos 16(3):155–191

Levy N (2019) No-platforming and higher-order evidence, or anti-anti-no-platforming. J Am Philos Assoc 5(4):487–502

Lewiński M (2022) Challenging authority with argumentation: the pragmatics of arguments *from* and *to* authority. Languages 7(3):207. https://doi.org/10.3390/languages7030207

Lewis D (1979) Scorekeeping in a language game. In: Bäuerle R, Egli U, Stechow A (eds) Semantics from different points of view. Springer, pp 172–187

Maitra I (2012) Subordinating speech. In: Maitra I, McGowan MK (eds) Speech and harm: controversies over free speech. Oxford University Press, Oxford, pp 94–120

Matsuda M (1993) Public response to racist speech: considering the victim's story. Words that wound: critical race theory, assaultive speech, and the first amendment. Westview Press, Boulder, pp 17–52

McClure E (2020) Escalating linguistic violence. In: Freeman L, Weekes Schroer J (eds) Microaggressions and Philosophy. Routledge, pp 121–145

McDonald L (2022) Reimagining illocutionary force. Philos Q 72(4):918–939

McGowan MK (2004) Conversational exercitives: something else we do with our words. Linguist Philos 27(1):93–111

McGowan MK (2009) Oppressive speech. Australas J Philos 87(3):389–407

McGowan MK (2019) Just words: on speech and hidden harm. Oxford University Press, Oxford

Peters U, Nottelmann N (2021) Weighing the costs: the epistemic dilemma of no-platforming. Synthese 199(3):7231–7253

Popa-Wyatt M, Wyatt JL (2018) Slurs, roles and power. Philos Stud 175(11):2879–2906

Raz J (1986) The morality of freedom. Clarendon Press

Raz J (2009) The authority of law: essays on authority and law. Oxford University Press, Oxford

Russell G (2019) Subordinating speech and speaking up. In: Lepore E, Sosa D (eds) Oxford studies in philosophy of language, vol 1. Oxford University Press, pp 178–207

Schiller H (2021) Is that a threat? Erkenntnis 86(5):1161–1183

Simpson RM (2021) Language and legitimation. In: Khoo J, Sterken R (eds) The Routledge handbook of social and political philosophy of language. Routledge, pp 345–361

Simpson RM, Srinivasan A (2018) No platforming. In: Lackey J (ed) Academic freedom. Oxford University Press, pp 186–209

Tirrell L (2012) Genocidal language games. In: Maitra I, McGowan MK (eds) Speech and harm: controversies over free speech. Oxford University Press, Oxford, pp 174–221

Tirrell L (2018) Authority and gender: flipping the F-switch. Fem Philos Q 4(3):1–32

Waldron J (2012) The harm in hate speech. Harvard University Press

Whitten S (2019) Recognition, authority relations, and rejecting hate speech. Ethical Theory Moral Pract 22(1):555–571

Witek M (2019) Accommodation in linguistic interaction: on the so-called triggering problem. In: Stalmaszczyk P (ed) Philosophical insights into pragmatics. De Gruyter, pp 163–191