“Offensiphobia” is a Red Herring: On the Problem of Censorship and Academic Freedom

Ben Cross¹ · Louise Richardson-Self²

Received: 29 November 2018 / Accepted: 30 September 2019
© The Author(s) 2021, corrected publication 2021

Abstract

In a recent article, J. Angelo Corlett criticises what he takes to be the ‘offensiphobic’ practices characteristic of many universities. The ‘offensiphobe’, according to Corlett, believes that offensive speech ought to be censored precisely because it offends. We argue that there are three serious problems with Corlett’s discussion. First, his criticism of ‘offensiphobia’ misrepresents the kinds of censorship practiced by universities; many universities may in some way censor speech which they regard as offensive, but this is seldom if ever a manifestation of ‘offensiphobia’. Second, we attempt to reconstruct Corlett’s criticism of ‘offensiphobia’ as a criticism of the practice of censoring hate speech, and show that this argument is unsuccessful. Third, we offer some brief reflections on how labelling universities as ‘offensiphobic’ is especially problematic in light of the current climate of political interference in university research and teaching.

Keywords Freedom of speech · Freedom of expression · Academic freedom · Censorship · Offense · Hate speech · Oppression

1 Introduction

There is no shortage of complaints that the once sacrosanct value of academic freedom is now being trampled on by the plagues of safe spaces and political correctness. The present generation of students, we are often told, is an excessively sensitive lot, more concerned with protecting themselves and each other from offense

This article is the result of an equal collaboration from the authors. Names are listed alphabetically.

✉ Ben Cross
bcro8137@uni.sydney.edu.au

Louise Richardson-Self
louise.richardsonself@utas.edu.au

¹ Wuhan University, Wuhan, China

² University of Tasmania, Hobart, Australia

than the importance of good scholarship, rigorous debate, and the pursuit of truth. J. Angelo Corlett goes so far as to describe this phenomenon as ‘offensiphobia’. As he sees things, ‘increasing numbers of college and university administrators, faculty, staff and students adopt the position that certain expressions ought to be censored because of their offensiveness to *them*’ (2018: 118, original emphasis). He subsequently levels a barrage of arguments against offensiphobia, concluding that it is a morally indefensible position.¹ We think there are at least three serious problems with Corlett’s critique.

The first of these problems concerns the accuracy of the description. Offensiphobia is not the same thing as expression intolerance. Let us grant that expression intolerance—a refusal to tolerate certain kinds of utterances, often characterized by some sort of censure—is increasingly common in higher educational institutions and that these utterances are often found highly offensive by many. It is this intolerance and these calls for censure to which Corlett objects. However, offensiphobia offers only a possible explanation for expression intolerance, and, as we argue, it is definitely not the most plausible one. The offensiphobe, as Corlett states, believes that certain utterances are worthy of censure *because of* their offensiveness. We argue, however, that few people in higher education, if any, are offensiphobes, and that expression intolerance can be easily explained without attributing offensiphobia to anyone. This is addressed in Sect. 2.

The second problem concerns the normative case which Corlett builds against offensiphobia. Since we believe that Corlett errs in his description of expression intolerance as a manifestation of offensiphobia, we also believe that much of his normative case attacks a straw man. Nonetheless, it is possible to extract certain aspects of his argument which instead target an existing form of expression intolerance—namely, calls to censure “hate speech” on university campuses. Though Corlett claims that hate speech ought not to be subject to censure, we show that his arguments for this view are also presently unconvincing. Our aim is not primarily to argue for the regulation of hate speech, but to show that Corlett’s case for the opposite view fails. This is addressed in Sects. 3 and 4.

Finally, by way of a conclusion, we suggest that the label of ‘offensiphobia’ is especially problematic in light of present developments in the politicization of higher education. Ironically, this is because it lends strength to many of the political forces which are currently seeking to *restrict* the freedom and integrity of university research and teaching. This is addressed in Sect. 5.

¹ Corlett focuses on the U.S. context, presumably because he believes offensiphobia is especially prevalent there. However, there are similar conversations around academic freedom in higher educational institutions in other Western countries; thus, it seems plausible to suppose that he would extend his criticism to these locations as well. It is true that the legal context of the U.S. differs somewhat from other countries, owing to the constitutional protections accorded to speech by the First Amendment. However, Because Corlett is at pains to emphasize that his critique of offensiphobia is based on *moral* grounds, not merely on the difficulties of implementing offensiphobic policies in a way that is consistent with the First Amendment, we believe he would apply this critique to institutions outside the U.S.

2 Offensiphobia is (Mostly) a Myth

Corlett defines offensiphobia as 'the fear of offensiveness and the attempt to prohibit it by way of law or public policy' (2018: 116). The basis for this claim is 'the presumption that people have a right to not be offended' (2018: 125), thereby 'implying a moral duty of others to not offend' (2018: 120). From this, we can see that the offensiphobe is committed to the following two propositions:

1. If the utterance of *X* offends any person *P*, then utterances of *X* ought to be censored in some way.
2. The brute fact that *X* offends *P* is what justifies the censoring of utterances of *X*.

Offensiphobia, so understood, is a possible explanation for acts of expression intolerance. Corlett, it seems, takes it for granted that offensiphobia is the only explanation of expression intolerance, even to the point of treating them as if they were the same thing.² For example, he states that, 'without a right to not be offended, those who seek to curtail higher educational academic freedom rights by way of censorship stand on unreasonable ethical grounds to do so' (2018: 113–114). However, if it is possible for something other than offensiphobia to explain expression intolerance, then a refutation of offensiphobia does not entail that all acts of expression intolerance lack a plausible justification. Still, one may think that offensiphobia offers the best explanation for expression intolerance as it currently exists in contemporary higher educational institutions, though we believe this is mistaken, as will shortly become clear.

What other explanations for expression intolerance are there? Here are two candidates, both of which may be easily confused with offensiphobia:

1. INDEPENDENCE: certain utterances or individuals ought to be censored because of reasons that are independent of the utterances and individuals' capacity to offend. They may frequently cause offense, but it is not the fact that they offend people that generates reasons for their censure. Here are three examples of views which defend certain practices of censure for certain kinds of speech:
 - i. Waldron (2012) argues racial slurs ought to be censored, not because they offend, but because they are a form of group libel that undermines the public good of mutual assurance concerning the fundamentals of justice.
 - ii. Simpson and Srinivasan (2018) argue that the practice of 'no-platforming' may, on some occasions, be justifiably used to censure speakers who lack the relevant disciplinary credentials, in order to defend the integrity of academic discourse.

² We acknowledge that Corlett thinks some categories of speech should not be legally protected and are not morally permissible. We will argue that he nonetheless attributes offensiphobic motives too freely to expression intolerance and does not provide compelling reasons to think the scope of presently unprotected speech is sufficient.

- iii. Finlayson (2012) argues that certain forms of censure may be an appropriate response to situations where a particular speaker monopolizes or dominates the space available for public discussion.
2. CONFIDENCE: this entails the acceptance of INDEPENDENCE, but with the additional claim that if certain people—especially historically oppressed constituencies—are offended by an utterance, and/or believe it to be worthy of censure, this *generally* reliably tracks the existence of good reasons for censuring it. Proponents of standpoint epistemology, for example, believe that ‘the different material conditions of people’s lives...affect what we are attuned to notice and what we are likely to see or experience;’ hence, it makes sense to hold that a marginalized group’s offense is grounded in good independent reasons even when a member of a dominant group does not automatically understand the expression as offensive (Richardson-Self 2018: 13).³

INDEPENDENCE and CONFIDENCE will often go together, although we can see that it is at least theoretically possible for them to come apart. A person can accept INDEPENDENCE without accepting CONFIDENCE, but not the other way around.

Offensiphobia differs from INDEPENDENCE and CONFIDENCE in two ways. The first main difference concerns their scope of application. Offensiphobia holds that an utterance ought to be censored if it offends anyone at all. The scope of INDEPENDENCE and CONFIDENCE is considerably less broad in this respect, since they suggest only that certain utterances which offend certain people should be subjected to expression intolerance in certain circumstances. However, INDEPENDENCE and CONFIDENCE have a broader scope in another respect: they allow for the possibility that certain kinds of expression intolerance may be appropriate even if the utterance or individual in question offends no one. Offensiphobia does not explicitly deny this, since it states that the causing of offense is a sufficient, rather than necessary condition for censure. However, this does mean that offensiphobia can only explain those forms of expression intolerance which call for the censure of something that offends.

Second, and most importantly, offensiphobia differs from INDEPENDENCE and CONFIDENCE in the way in which it locates the wrongness of giving offense. Offensiphobia holds that it is the fact that offensive utterances offend that makes it wrong to utter them. Both INDEPENDENCE and CONFIDENCE, on the other hand, do not hold that there is anything wrong with causing offense in and of itself, but hold that there are certain utterances which may commonly cause offense, for which there are reasons independent of their offensiveness explaining why such utterances ought to be censored.⁴

³ We are not here claiming that such a position is justified. Some may worry that CONFIDENCE overlooks the possibility that oppressed groups may suffer from a ‘false consciousness’ which may diminish the epistemic reliability of their views. Others, like standpoint epistemologists, may defend CONFIDENCE by claiming that the oppressed can have a ‘double-consciousness’ which enhances their epistemic reliability. Our point here is not that either view is correct. Rather, it is that INDEPENDENCE and CONFIDENCE offer alternative explanations for acts of expression intolerance.

⁴ This point about the difference between offensiphobia and INDEPENDENCE/CONFIDENCE parallels the distinction between ‘offense’ and ‘harm’ made in Waldron (2012: chapter 5) and McGowan (2018). In McGowan’s words, ‘offense is a subjective inner feeling; harm is an objective state’ (2018: 136). How-

Confidence makes the additional claim that, when certain people are offended by an utterance, the fact that they are offended is a reliable indicator of the existence of such reasons. In such circumstances, it may easily be mistaken for offensiphobia. The important difference, however, is that CONFIDENCE entails the belief that offense 'tracks' the existence of reasons for censoring an utterance, whereas offensiphobia entails that offense 'makes' or constitutes the reason itself.

Thus, it is easy to see how the existence of expression intolerance can result from widespread acceptance of INDEPENDENCE and CONFIDENCE. If the belief that certain utterances and individuals ought to be censored is more widespread today, this may be because many people now believe that such utterances can be harmful, dangerous hindrances to participation in academic discussion, causes of distorting effects on academic research, or problematic in any other way that warrants censure. And if, in some circumstances, people appeal to the mere fact that a particular person has been offended by an utterance as grounds for censure without, let's say, explicitly weighing these sorts of considerations, this can be explained by CONFIDENCE. We certainly do not claim that INDEPENDENCE and CONFIDENCE actually *justify* expression intolerance. Justification depends on the quality of the reasons. Our point here is simply that expression intolerance may be explained by INDEPENDENCE and CONFIDENCE, rather than offensiphobia.

But we also believe that this explanation is significantly better than offensiphobia. One obvious reason for this has already been raised: there are forms of expression intolerance that target utterances and individuals, though they do not appear to offend anyone. Think of a protest against a pro-war politician invited to speak at a university campus.⁵ Or, to use a prominent contemporary example, consider the 'UnKoch my campus' campaign, which protests what it sees as the corrupting effect of university donations by the Charles Koch Foundation.⁶ It may even be the case that certain kinds of language which do not normally offend anyone may nonetheless contribute to a climate that is hostile towards a particular group (Baker 2012: 70).⁷ Some may think that such speech ought to be censored, even in the absence of offense.

Furthermore, we believe that INDEPENDENCE and CONFIDENCE offer a much better explanation for expression intolerance than offensiphobia even when the utterance or individual in question *does* offend someone. Let us demonstrate this by examining a kind of expression intolerance commonly found in higher educational institutions today: the person who calls for the censure of utterances she regards as racist.

Footnote 4 (continued)

ever, there is an important difference. INDEPENDENCE and CONFIDENCE do not limit the range of possible reasons for censure to the harm resulting from the violation of dignity, as Waldron does, or even to the category of harm at all. For example, Simpson and Srinivasan (2018: 190–194) argue that the liberal case for no-platforming does not rest on Mill's harm principle. Our point is not that reasons for censure which are not concerned primarily with harm are *good* reasons (or bad, for that matter), but only that INDEPENDENCE and CONFIDENCE allow that such reasons may exist.

⁵ See for example Herszenhorn (2006).

⁶ See <http://www.unkochmycampus.org/>.

⁷ Note that Baker, like Corlett, does not advocate for the legal censoring of hate speech.

Call her the *anti-racist activist*.⁸ What form would the anti-racist activist's beliefs and actions take if she accepts INDEPENDENCE or CONFIDENCE, and how does this contrast with the form they would take if she accepts offensiphobia?

We think it will become obvious that the former offers a much more realistic explanation. Hence, we think that the anti-racist activist who accepts INDEPENDENCE and CONFIDENCE is real, common, and a key contributor to existing instances of expression intolerance, whereas the anti-racist activist who accepts offensiphobia can only be, at best, a minor player. Our argument consists of two key premises. The first is that the anti-racist activist who accepts INDEPENDENCE or CONFIDENCE can take a position that is consistent with her actions and expressed beliefs, whereas the anti-racist activist who accepts offensiphobia takes a position that is so obviously inconsistent and implausible that it is hard to see how anyone might genuinely do so. To see how this is the case, let us consider three points of comparison.

1. If the anti-racist activist accepts INDEPENDENCE, she will say her call for the censure of racist speech stems from the wrongness of racism. She need not deny that racist speech frequently offends people, especially certain historically oppressed constituencies. But she will add that it is the wrongness of racism that justifies the censure of racist speech, and not the brute fact that it offends such people. If she accepts CONFIDENCE, she thinks the fact that certain utterances offend a particular historically vilified group tracks *very good reasons* for regarding these utterances as wrongful *and* worthy of censure. But what if the anti-racist activist was actually an offensiphobe? She would claim that racist speech is worthy of censure purely because it is offensive. If she thinks that there is anything else that is wrong with racism, she does not believe that this has anything to do with why racist speech ought to be censored. Instead, her sole concern is with the prevention of offense. This much demonstrates how the anti-racist activist who accepts INDEPENDENCE and CONFIDENCE holds beliefs that are vastly different to the anti-racist who accepts offensiphobia. The former believes there is a fundamental connection between the wrongness of racism and the appropriateness of censuring what she believes to be racist speech, and that it is the existence of this connection which justifies her position. The offensiphobe, on the other hand, denies that this connection exists.
2. A more practical difference becomes clear when the anti-racist activist is confronted with racists who are offended by *anti-racist* speech. If she accepts INDEPENDENCE, she does not believe there is anything about offensive speech which is necessarily censure-worthy; hence, she will not conclude that one may censure *anti-racist* speech. She will dismiss the racist's offense because they are, well, racist. But if she accepts offensiphobia, she finds herself in a dilemma: her anti-racist convictions (assuming that she has some such convictions) give her reasons to criticize racist attitudes, but her commitment to offensiphobia entails a duty not to say anything that will offend anyone, including racists. The offensiphobe,

⁸ We focus on the anti-racist activist for ease of presentation. But we assume that what we say about the anti-racist activist can also be said for other activists who call for the censure of other kinds of speech which commonly offend (e.g. sexist, homophobic).

it seems, must avoid virtually all forms of public speech or else be charged with holding double standards, since, as Corlett rightly points out, virtually all utterances offend someone (2018: 125).

3. Finally, consider how the anti-racist activist deals with the possibility of mistaken beliefs. The anti-racist activist who accepts INDEPENDENCE may or may not be correct in her belief that she has *good* independent reasons to call for the censoring of speech she regards as racist. Perhaps she frequently misidentifies speech as racist. Or perhaps she is mistaken in her belief that the wrongness of racism justifies the censure of racist speech. But it will matter very much to her that she is not mistaken. If she could be convinced that she has misidentified a particular utterance as racist, even if she found it offensive, we would plausibly expect her to revise her judgment that it ought to be censored. Yet we would expect the anti-racist offensiphobe to be largely indifferent to the possibility that she has misidentified speech as racist. Suppose she has long considered the utterance of *X* to be racist and found it to be highly offensive. If she is then persuaded that *X* is not, after all, a form of racist speech, she would then say 'I still found *X* offensive, even though it wasn't racist! The fact that I found it offensive is sufficient to justify its censure.' Of course, in learning that *X* is not a racist utterance, it may be that she no longer finds it offensive. But there is no guarantee. And if she remains offended, she will still believe the speech ought to be censored. This, we think, makes it clear that she is not really an anti-racist activist after all.

These three distinctions are summarised in Table 1 below.

It seems unlikely that the positions described in the right column bear any resemblance to actual existing anti-racist activists. For certain groups who have a more or less identifiable *raison d'être*, it is easy to show how this is the case. Consider Antifa, which in recent years has acquired a reputation for using confrontational and occasionally violent tactics to combat racist speech. Antifa regards racist speech as censure-worthy because it considers racism to be a manifestation of fascist politics, and holds that fascist politics requires forceful opposition (Bray 2017). It does not regard the fact that racist speech offends people as in any way relevant, nor, for that matter, is it troubled by the fact that its own actions may offend its opponents.

When it comes to determining which column each individual activist may belong to, however, things are somewhat trickier. Short of interviewing each one, it is

Table 1 Possible explanations for the anti-racist activist's expression intolerance

INDEPENDENCE/CONFIDENCE	Offensiphobia
Racist speech (e.g.) is wrong, ∴ censure racist speech	Racism (e.g.) offends, ∴ censure racist speech
Offensive speech is permissible if there are good independent reasons	Offensive speech is never permissible; one has a right not to be offended
If judgments about the racism (e.g.) of utterances are mistaken, attitudes towards censure will change	The correctness of judgments about the racism (e.g.) of utterances is irrelevant; only the presence of offense matters

difficult to see how this can be known with complete certainty. We will, however, conjecture that the following two claims are correct: no anti-racist activist will (a) consciously endorse the content of the right column (can you imagine a self-proclaimed ‘anti-racist activist’ who does not believe that the wrongness of racism has anything to do with the need to censure racist speech?), and (b) act in a way that is consistent with the right column’s requirements. (Those who protested Richard Spencer’s presence at Florida University in October 2017 did not return the next day to protest lectures on Critical Race Studies.) The anti-racist activist, then, can accept offensiphobia *only at the cost of extreme inconsistency with her own stated beliefs and actions*. On the other hand, she can accept INDEPENDENCE and CONFIDENCE without any such inconsistency. This is not to say that her expression intolerance is justified; her reasons for expression intolerance may or may not be good ones. Our point is only that her conscious beliefs and actions will fit neatly with the content of the left column.

Our second premise is that, other things being equal, we should prefer an explanation of a person’s behavior that does not attribute wildly implausible positions to them over one that does. The critic of offensiphobia holds that it is an utterly indefensible position, incompatible with the basic normative commitments of most people, and should be wholeheartedly rejected. But given what we have seen so far, it more clearly resembles a pathological confusion, bordering on neurosis, than a genuine (and well-reasoned) position. Of course, the fact that a position is indefensible does *not* always prevent people from endorsing it, least of all in academia! But if we wish to attribute such a position to anyone, we should at least provide a great deal of evidence for why other more plausible positions do not accurately explain their behavior.

Corlett offers no such explanation. In fact, his entire treatment contains only two real-world examples, neither of which can be plausibly construed as offensiphobia. The first example is a 2011 “Dear Colleague” email from the Secretary of Defense, threatening to withdraw funding from institutions which fail to comply with various mandates, which, according to Corlett ‘have had a deleterious effect on academic freedom for faculty’ (2018: 115–116n2). Given that the letter to which he refers does not purport to ban a list of words, but focuses instead on the problem of ‘harassment’—a form of speech which Corlett acknowledges should not be protected (2018: 122)—it is hard to see how this is in any way connected to offensiphobia. The second example he raises is that of the Industrial Workers of the World, where he repeats the propaganda of American newspapers in 1912, that the IWW were ‘hypocrites’ for demanding the right to denounce capitalism ‘while simultaneously advocating and even seeking to violate the same rights of others’ (2018: 117n6). Whatever one thinks of the IWW’s attempts to control internal dissent, its actions were clearly motivated by strategic political concerns that had nothing remotely to do with offense (Rabban 1994: 1078–1080).⁹ To say that Corlett falls short of

⁹ Curiously, Corlett does not see fit to mention that many of the IWW’s opponents did attempt to deny its right to speak on the grounds that its speeches were an ‘offense to common decency’, and in so doing may have themselves come close to a form of offensiphobia (Rabban 1994: 1094).

providing the evidence necessary to make offensiphobia the best explanation of expression intolerance would be a significant understatement.

So, to sum up, we have argued that INDEPENDENCE and CONFIDENCE offer a much better explanation for existing forms of expression intolerance than offensiphobia. Our argument for this claim has consisted of two premises: first, that the anti-racist activist's expression intolerance can be explained by offensiphobia only by attributing to her a severely inconsistent set of theoretical and practical positions, whereas it can be explained by INDEPENDENCE and CONFIDENCE without such inconsistencies; and second, we should, other things being equal, prefer an explanation for a person's behavior which does not commit her to a radically implausible position over one which does. Hence, we think that INDEPENDENCE and CONFIDENCE provide the superior explanation for the expression intolerance of anti-racists, and by extension, that of activists in service of other causes. Genuine offensiphobes, if they exist at all, are few and far between, and have had little if any role in creating the kind of culture in higher educational institutions of which Corlett is so critical.

3 Implications for Corlett's Arguments

If existing expression intolerance is best explained by acceptance of INDEPENDENCE and CONFIDENCE rather than offensiphobia, then it follows that Corlett's arguments against expression intolerance attack a straw man. In this section, we examine four of Corlett's arguments—(1) the subjectivity of offense makes offensiphobia impossible to implement; (2) offensiphobia is a form of moral hypocrisy; (3) offensiphobia relies on special pleading; and, (4) offensiphobia has paternalistic implications—and show how this is the case.

1. Corlett has a great deal to say about the problem posed by the 'subjectivity' of offensiveness (2018: 116–117, 119, 121, 124–127, 129, 135–142). Roughly put, if an utterance is offensive to some, but not others, then it is not clear whether it should be properly classed as 'offensive' and censured accordingly. Offensiphobia, Corlett claims, is therefore incapable of being implemented. But the problems posed by the subjectivity of offensiveness melt away when we realize that virtually no one is calling for the censure of all utterances which offend someone. Instead they hold, in accordance with INDEPENDENCE or CONFIDENCE, that there are reasons for censuring some offensive utterances and not others. Many of those who defend some forms of expression intolerance do so, for example, on the basis that certain utterances do not merely offend, but instead *harm*.¹⁰ If correct, the subjectivity of offense does not pose any problem to the implementation of expression intolerance.

¹⁰ See our discussion in Sect. 4. Corlett, for his part, acknowledges that certain kinds of speech that are unprotected by the First Amendment are indeed harmful, although, as we will see in the next section, his claim that these are the only kinds of genuinely harmful speech is unjustified.

2. Corlett also accuses proponents of offensiphobia of moral hypocrisy (2018: 117). He cites what he regards as left-wing manifestations of offensiphobia, where people support flag desecration but call the censure of racist language. He then places this alongside what he regards as right-wing manifestations of offensiphobia, where people support the use of words which others regard as racist, but call for the censure of flag desecration. Both sides, he argues, deploy offensiphobia in a way that applies one standard for themselves and another standard for others. But if the expression intolerance of both sides is explained by INDEPENDENCE or CONFIDENCE, then either side will claim that there is an important asymmetry between their side and the other: they have good reasons for censuring what they find offensive; the other side does not. Of course, at least one side must be wrong about this, but if they believe that such an asymmetry exists, their brand of expression intolerance is not a form of moral hypocrisy.¹¹
3. Corlett further argues that offensiphobia commits the fallacy of ‘special pleading’ (2018: 136). The offensiphobe, he claims, appeals to the interests of a particular group in order to gain an arbitrary exemption from the rule of freedom of expression. But if expression intolerance is defended on the grounds of INDEPENDENCE or CONFIDENCE, then its proponent does not consider this exemption as arbitrary, but based on good reasons. Again, she may be mistaken about this, but if this is so, her expression intolerance is mistaken for reasons other than special pleading.
4. Finally, Corlett criticizes offensiphobia for its paternalistic implications (2018: 127–130). His main concern seems to be how certain linguistic norms, ostensibly developed to treat oppressed constituencies with greater respect, might nonetheless come into conflict with the preferred language of some people *among* these constituencies. This may be because they have sought to reclaim or re-appropriate formerly derogatory terms for themselves (e.g. women’s use of the word ‘bitch’), or because they reject what they perceive as the external imposition of a label (e.g. Indigenous peoples’ opposition to the term ‘Native American’). As Corlett rightly points out, we can further distinguish the way in which a racial epithet

¹¹ One might worry that if debates about the appropriateness of censure are to be understood in this way, this removes any possibility of a politically ‘neutral’ position. Perhaps one appealing feature of ‘offensiphobia’, among its many other shortcomings, is that it does purport to establish such a neutral position—that which is offensive to the political right is just as censure-worthy as that which is offensive to the political left. If we abandon any hope of such a ‘neutral’ position, does this mean we are condemned to a disastrous situation where the left and right, rather than trying to achieve compromise or consensus, are instead only trying to censure each other? While there may be many reasons to be sceptical of the possibility or desirability of such ‘neutrality’, we think the most important response to this worry is to emphasise how complex and contextual good reasons for censoring speech may need to be. Suppose, for example, the political left believes that much of the speech of the right is genuinely harmful. It still might hold that, in many cases, censoring harmful right-wing speech could have even more harmful consequences. For example, it might drive it underground, or create a precedent for the political right to exploit when they have the opportunity to do so. On this point, see especially Harvey (2012). The point is that, if the prospect of allowing expression intolerance to become a political football is a reason for refraining from expression intolerance in some circumstances, it is a reason that both INDEPENDENCE and CONFIDENCE can accommodate.

might be used by some group members as a way of denoting affection, from a plainly derogatory use by (e.g.) white supremacists. He refers to the former usage as 'cacophemistic' and the latter as 'disphemistic' (2018: 128).¹² In such cases, Corlett argues, those who are offended by violations of these linguistic norms and seek to censor their violation thereby adopt a paternalistic attitude towards the very people whose interests they purport to defend.

We happily grant that paternalism is normally a bad idea, especially when directed towards oppressed peoples. But INDEPENDENCE does not encourage paternalism. If we accept that there are no good reasons for subjecting a constituency's reclamation of a term to any kind of censure, then it follows that INDEPENDENCE requires that the constituency's use of the term be left uncensored. It may still be the case that the term should be censored in some contexts, or when used by people outside the constituency. Yet, if we can recognize these distinctions, there is no reason why university administrators can't recognize them too. There doesn't seem to be anything incoherent or *necessarily* paternalistic in holding that the use of certain terms ought to be censored in some contexts but not others. However, one might think that the existence of diverse views *among* oppressed constituencies poses a problem for the non-paternalistic implementation of CONFIDENCE. This becomes especially clear when we note that some members of a particular constituency find a certain word or phrase offensive, while others do not. If this is the case, we may think that CONFIDENCE can be implemented only in a paternalistic way, or not at all.

Consider an example. Some women have reappropriated the term 'slut', while other women remain uncomfortable with this term and see it as a slur (Reger 2015). If some women use the word 'slut' in a way that offends other women, this does seem to place the proponent of CONFIDENCE in a difficult position. But this difficulty seems insurmountable only if we make the mistake of construing CONFIDENCE in universalist, absolutist terms. The proponent of CONFIDENCE may say that the appropriate response will depend a great deal on the context and the full details of the speaker's utterance. She might say that the best approach for university administrators is to encourage greater caution and sensitivity as opposed to implementing a more plainly coercive response. However, the complexities of these situations do not threaten the key point behind CONFIDENCE: if we want to find out whether there are good reasons for censoring terms laden with derogatory heritage, the responses of the very people to whom the term has historically been applied will *normally* give us a pretty good clue. Complex cases such as this simply necessitate a more nuanced application of this principle. More importantly, the existence of complex cases does not eliminate the existence of simpler ones. The fact that we might be uncertain about the status of certain cacophemistic uses does not mean we must also be uncertain about the status of plainly disphemistic uses.

¹² Additionally, we can distinguish between its cacophemistic use among group members and the misguided attempts of out-group individuals to use it in a similar fashion.

4 The Normative Case Against Expression Intolerance: Should Hate Speech be Censured?

We have thus seen that four of Corlett's arguments do not provide any critical purchase against existing expression intolerance, since they are best explained by acceptance of INDEPENDENCE and CONFIDENCE. But it may be that the argument which Corlett attempts to build against expression intolerance in tertiary institutions can, at least to some extent, be separated from his focus on offensiphobia. In this section, we have two aims. First, we will try to reconstruct Corlett's argument in such a way that does not direct it against offensiphobia, and instead focuses it on attempts to censure hate speech that are not motivated by offensiphobia. Second, we will try to show that this argument, as Corlett presents it, is unconvincing.

4.1 Corlett's Argument Reconstructed

Let us first clarify the position on hate speech that Corlett seeks to defend. One clear aspect of his position is that he is hostile to attempts to censure hate speech as a category of utterances. "Hate speech," he claims, 'is and should be protected expression under established U.S. law and solid college and university policies which are consonant with said law' (2018: 141). Corlett's argument for this position is not based on any kind of free speech absolutism, according to which expression intolerance is always unjustified.¹³ Instead, he defends what might be called a 'First Amendment' position on the use of censure. That is, an utterance is an appropriate candidate for censure if and only if it falls under one of the categories of speech not protected by the First Amendment. Corlett here follows Feinberg (1980 and 2004) in four ways:

1. Restricting freedom of expression is permissible only when doing so prevents 'direct and substantial harm to assignable persons' (2018: 114).
2. Direct and substantial harms are those which involve 'the setting back of a legitimate interest' (2018: 120, 123fn. 17, 124, 135, 141–143).
3. Utterances which involve the setting back of a legitimate interest are those which fall under one of the categories of speech not protected by the First Amendment (call them "the unprotected categories") (2018:121).
4. There are five unprotected categories: knowingly and unnecessarily causing panic; provoking retaliatory violence; inciting a violent insurrection; employing 'fighting words'; and violating another's privacy (2018: 121).

To reiterate: Corlett's reason for keeping expression intolerance confined to utterances covered by the unprotected categories is that only such utterances are harmful (see also Corlett and Francescotti 2002), and that the harmfulness of utterances

¹³ In fact, free speech absolutism is a position that few (if any) people endorse, as Maitra and McGowan (2012) point out.

is the only good reason for censure. Corlett does admit that some hate speech can harm, but he seems to believe that (in something of a happy coincidence) this harmful hate speech is already covered by the unprotected categories. For example, an utterance of the phrase 'white trash!' with the intention of provoking retaliatory violence may well be justifiably subject to censure (2018: 143). While forms of hate speech that are not covered by the unprotected categories can offend or cause "hurt feelings", they are not a setback of legitimate interests. This is because there is no 'legitimate interest in not being gratuitously offended' (2018: 142).¹⁴ So censoring hate speech would be both pointless (since harmful hate speech is already unprotected) and unjustified (since not all hate speech is harmful). Thus, its categorical censure would be an unjustifiable violation of freedom of expression.

We can therefore summarize Corlett's argument as follows:

1. Acts of censure are unjustified unless the utterance is harmful.
2. The only kinds of harmful utterances are covered by the unprotected categories.
3. Not all forms of hate speech are covered by the unprotected categories.
4. Therefore, attempts to categorically censure hate speech are unjustified.

Since many higher educational institutions do attempt to censure hate speech, this argument evades the strawman fallacy discussed in Sect. 3. Hence, we take this to be the most charitable reconstruction of Corlett's argument against offensiphobia. However, we also find this argument against the censure of hate speech wanting. In what follows, we focus exclusively on the second premise. We do not argue that it is necessarily false, but only that Corlett gives us no good reasons for thinking that it is true. However, our focus on the second premise should not be taken as a concession that the first premise is true, still less that Corlett provides good reasons for taking it as true.¹⁵ But since disagreements regarding the first premise are likely to be somewhat deeper and more intractable, we think it makes sense to focus on the second premise instead.

It seems Corlett has two reasons for thinking that the hate speech is harmful only when it falls under the unprotected categories. First, he employs a *reductio ad absurdum* against the view that 'hate speech', as a category, is harmful. Surely, he suggests, instances of hate speech like 'these Republicans can fuck-off!' are neither harmful nor censure-worthy (2018: 141–142). Hence, anyone who claims that hate speech is harmful *qua* hate speech is committed to the absurdity of regarding

¹⁴ Several scholars have characterized 'hate speech' as speech which offends (Archard 2014; Bannister 2008; Greene 2012; Schauer 2012), and several thinkers have opposed its regulation or ban for this very reason (see, for instance, Dworkin 2009; Erjavec and Kovičič 2012; George 2014; Haraszti 2012; Schulzke 2016). However, others who support hate speech bans tend *not* to characterize hate speech as merely offensive. As we will see, this failure to engage with the claims of pro-censure scholars is highly problematic.

¹⁵ For example, one might question whether genuine harm provides the only plausible grounds for acts of censure (Simpson and Srinivasan 2018: 194–195), or even whether the concept of "freedom of expression" is properly coherent at all (Finlayson 2012).

plainly non-harmful utterances as harmful (where harm is taken to involve a setback to legitimate interests). In order to make this *reductio* work, Corlett needs to define hate speech in such a way that includes these sorts of plainly non-harmful utterances.

Second, Corlett (2018: 142) argues that the offensiphobe has the ‘sole burden of argument’ when it comes to showing that offensive speech is harmful. Although, as we have seen, this focus on offensiphobia commits the strawman fallacy, perhaps Corlett would still argue that those who reject the second premise—those who believe that hate speech can harm in ways not covered by the unprotected categories—have the burden of proof. They must demonstrate conclusively that hate speech, as a category of utterances, is harmful. And Corlett seemingly thinks they have failed to do this. In what follows, we criticize both reasons.

4.2 First Argument: the *Reductio*

As mentioned, the *reductio* relies on Corlett’s definition of hate speech—but the claim that “anyone who views hate speech as harmful *qua* hate speech is committed to the absurdity of regarding plainly non-harmful utterances as harmful” is plausible only if we already understand hate speech in such a way that includes plainly non-harmful utterances. We will criticize this argument by criticizing Corlett’s definition of hate speech.

It seems Corlett relies on a definition of hate speech developed in an earlier work, according to which ‘hate speech’ is ‘symbolism, linguistic or otherwise, that expresses antipathy toward some group or toward an individual on the basis of membership in some group’ (Corlett and Francescotti 2002: 1083). Two aspects of this definition are noteworthy for our purposes here: (i) the scope of this definition is not confined to groups who are oppressed ‘in regular and significant ways’, since to say otherwise would be to neglect forms of oppression that are yet to occur (Corlett and Francescotti 2002: 1082), and (ii) it accords a significant role to intention: ‘whether a sentence has hateful content ultimately depends on the typical intentions of those who use that sentence’ (Corlett and Francescotti 2002: 1086).

This definition of hate speech is radically at odds with the term’s original meaning. Much like “sexual harassment” (cf. Fricker 2007), “hate speech” emerged as a way of making sense of *actual* experiences of linguistic violence in a context of systemic group-based prejudice—specifically, non-white racial prejudice (see Brown 2017a, b; Matsuda 1989; Richardson-Self 2018).¹⁶ The conceptual innovation involved in creating the concept of hate speech, occurring around the late 1980s (Brown 2017a: 424), was motivated by the belief that certain kinds of speech that referred to oppressed constituencies in a derogatory manner may have a distinctive kind of impact upon them.

¹⁶ It is worth noting that, even though Masuda (1989: 2131–2132) confines her analysis to non-white racial prejudice, she allows that the concept of ‘hate speech’, as she conceptualises it, could also be applied to speech targeting women and LGBTI people.

This 'original' definition of hate speech evidently differs significantly from Corlett's definition, most importantly for our purposes in the way in which it does not possess either of the two features we highlighted above. The original definition applies the label 'hate speech' only to forms of speech directed towards oppressed constituencies, because those who developed the concept held that a context of oppression makes a difference to the kind of impact that utterances can have (Gelber 2019). Similarly, it does not place much importance on the intentions of a speaker.¹⁷ A speaker can have hateful intentions when her utterance does not target oppressed constituencies, so *hateful* speech is not the same as *hate speech*. A speaker may also lack hateful intentions even though her utterance does target oppressed constituencies. Racism, as Matsuda (1989: 2332) acknowledges, can be unconscious as well as conscious.

Now the fact that the original meaning of hate speech differs from Corlett's definition doesn't automatically mean that the latter is defective. However, we think there are at least two good reasons for preferring the former over the latter. The first reason is that, in the context of a debate about hate speech and harm, Corlett's account allows him to claim a definitional victory, while ignoring the real issue at hand. What matters, for the purposes of this debate, is whether utterances targeting oppressed constituencies are harmful in ways not covered by the unprotected categories. If Corlett's definition were accepted, it would not so much settle the debate as shift it: instead of talking about whether hate speech is harmful, we would instead talk about whether utterances targeting oppressed constituencies are harmful. The fact that Corlett's definition of hate speech includes plainly non-harmful utterances which do not target oppressed constituencies does not help us address this question.

Second, as we have seen, the original definition of hate speech was developed in order to meet a specific epistemic need, namely, to recognize the distinctive impact of utterances targeting oppressed groups. This original definition conceptualizes hate speech as a "thick" concept which both describes a set of utterances and applies a negative evaluative status to them, thereby giving them a different evaluative status to utterances that do not target oppressed groups. If Corlett's definition were widely adopted, "hate speech" could no longer serve this purpose. It wrongly implies that "those Republicans can fuck-off!" and "those bitches can fuck-off!" are evaluatively equivalent, at least until we can invoke some further evaluative concept to distinguish between them. This conflation of hateful speech and hate speech could be a way of "inflating" the concept of hate speech, so that utterances like "those Republicans can fuck-off" come to be seen as no less objectionable than utterances targeting oppressed groups.¹⁸ Or, perhaps more likely, it would "deflate" the concept, so that hate speech loses its negative evaluative status and both kinds of utterances (*hateful* speech and *hate speech*) come to be seen as merely offensive, and not seriously objectionable.

To demonstrate that this conflation occurs in "Offensiphobia", consider Corlett's posing of the following rhetorical question: 'Even if a causal link between hateful

¹⁷ On this point, see especially Brown (2017a) on "The Myth of Hate".

¹⁸ Note that even though Corlett's definition would 'inflate' the concept of hate speech in this respect, it would also shrink its scope in other ways: it would deny the status of 'hate speech' to utterances targeting oppressed groups that are not uttered with hateful intent (e.g. homophobic slurs uttered out of ignorance).

expressions and genuine harms could be established’—which implies Corlett thinks such a link cannot be established—

...would it not include hatefully offensive expressions such as “Those god-damn ignorant Christians need to wake up and see that there is no God!”, or “Those fucking creationists can go to hell!” or “Those Republicans can fuck-off!” or “Trump is not my mother-fucking President!” or any number of other opinions which are uttered by many leftist academics throughout the U.S.? (2018: 141–142, our emphasis).

In either the “inflationary” or “deflationary” scenario, the concept “hate speech” can no longer be used to distinguish between the evaluative status of utterances targeting oppressed groups and utterances that do not. This would return us to the position of those who were formulating the concept of hate speech in the 1980s, needing a new concept to make sense of the special evaluative importance of utterances which target oppressed peoples. Again, if a concept is to serve this purpose, it must not be extended to utterances that do not target oppressed peoples, while including all utterances that do—regardless of the speaker’s intentions. At some point, this regress needs to end, and the concept of hate speech seems the most appropriate place to end it.

One might object that the fact that Corlett’s definition does not serve the original purpose of hate speech does not necessarily count against it. We can make two points in reply to this objection.

First, there are good reasons for thinking that the original purpose is a laudable one. Utterances which do not target oppressed constituencies operate in a different way to utterances that do. Consider Miranda Fricker’s idea of ‘tracker prejudices’: utterances which target oppressed constituencies “‘track’ the subject through different dimensions of social activity—economic, educational, professional, sexual, legal, political, religious, and so on” (Fricker 2007: 27). Such utterances impact their targets in ways that other utterances do not, and often do so regardless of the speaker’s intentions. This isn’t the same thing as claiming that utterances targeting oppressed groups are harmful—although those who developed the concept of hate speech evidently thought so—but only that the existence of a context of oppression makes a difference to what utterances do, and that this difference should be reflected in how we evaluate them.

Second, Corlett’s redefinition of ‘hate speech’ away from the original concept seems a clear case of what Emmalon Davis (2018) calls ‘epistemic appropriation’. This occurs when a concept developed by marginalized peoples gains wider recognition, but in a way that is ‘detached’ from the marginalized peoples responsible for its innovation. It may also lead to the benefits of this concept being ‘misdirected’ away from marginalized peoples and towards the powerful. Epistemic appropriation, so understood, is harmful because it estranges marginalized peoples from their own epistemic capabilities (Davis 2018: 723–724). Again, this is not to assert that hate speech is harmful, but that insofar as Corlett’s definition of hate speech amounts to epistemic appropriation, *this definition itself* is harmful.

We should therefore accept that ‘hate speech’, best understood, applies only to utterances targeting oppressed constituencies, and to all such utterances regardless

of the speaker's intentions. If this is correct, then the claim that hate speech can harm in ways unrecognized by the unprotected categories does not lead to the absurdities that Corlett believes it does. This is because, as we have seen, plainly non-harmful utterances like "these Republicans can fuck-off!" are not hate speech, no matter how hateful the intent of the speaker may be. Corlett's *reductio* against the claim that hate speech is harmful in ways not included in the unprotected categories is therefore unsuccessful.

4.3 Second Argument: A Burden of Proof?

Let us now turn to Corlett's second argument. This argument holds that those who claim that hate speech is harmful in ways not covered by the unprotected categories have the "burden of proof" and that, at least so far, they have not met this burden. This is most evident in passages such as the following:

Using the locution "hate speech" to describe certain expressions which one finds ought to be censured by law appears to be little more than a sleight of hand to use the language of offensiphobia in order to bully one's way philosophically through a complex problem (Corlett 2018: 140).

And:

"Legitimate" is important here, as it appears that the concern begs the question of whether or not there is a legitimate interest or right to not be offended or to have certain words not expressed about either themselves or others.... while the expression of hate might in many cases lead to the feeling of being harmed by many, it is far from clear that there is a causal link between hateful expressions, on the one hand, and genuine harms, on the other. Hurt feelings caused by hateful expressions do not amount to the causing of genuine harms in the requisite sense (Corlett 2018: 141).

It may not always be inappropriate to assert that the burden of proof lies with one side rather than the other. Typically, this sort of claim is most fitting when one side has assumed what turns out to be a crucial yet contestable premise, without arguing for it. This seems to be what Corlett has in mind when he claims that the view that hate speech should be censured on university campuses 'assumes that certain expressions constitute genuine harms when it is unclear that they do' (2018: 142). It may well be true that an offensiphobia-motivated demand for the censure of hate speech would make this assumption. But the philosophical literature on this topic does not make this assumption. Waldron (2012), Langton (2009, 2012), Maitra and McGowan (2012), and West (2012), among many others, have all argued that hate speech has certain harmful capacities. So, at first glance, Corlett's claim that the burden of proof lies on those who think hate speech is categorically harmful appears seriously out of place. The harmfulness of hate speech has not simply been assumed;

it has been argued for. Why, then, might Corlett still think it appropriate to claim that the burden of proof lies with them, short of unfamiliarity with the literature?

One possibility is that he is aware that arguments have been made for view that hate speech is harmful but thinks that they are in some way circular, and have basically assumed what they need to argue for despite the appearance of having done otherwise. Again, it is not uncommon for philosophers to claim that arguments do this, but such claims are appropriate only after close scrutiny of the literature in question. Of the authors listed above who argue that hate speech is harmful, Corlett (2018: 139–140) cites only one (Waldron), and claims his argument commits the fallacy of special pleading. This criticism, we have seen, relies on the misleading view that Waldron's argument is based on offensiphobia, and so misfires. Although he does acknowledge that Waldron distinguishes between offense and harm, Corlett then resorts to claiming that it faces the *reductio* discussed above, and that it assumes what it needs to argue for (2018: 140–142). We have already seen that the *reductio* fails. The subsequent assertion that Waldron assumes what he needs to argue for, meanwhile, comes without any further textual evidence or discussion of the content of Waldron's argument. So far, then, Corlett's claim that those who believe hate speech is harmful have the burden of proof seems seriously inapt.

Is there another explanation for this mistaken analysis? We think there is a latent assumption underscoring Corlett's misreading of Waldron that would also explain his dismissal of the view that hate speech can be categorically harmful. This has to do with the distinction between causal and constitutive harms — a distinction that Corlett fails to acknowledge. A clear example of the former is John Stuart Mill's (1978: 53) famous example of a person who shouts, "corn dealers are starving the poor!" to an angry mob who have assembled outside the corn dealer's house.¹⁹ This utterance causes harm by inciting violence, and so falls under the unprotected category of provoking retaliatory violence. Utterances that constitute harm, on the other hand, are harmful in the very act of their being uttered. Consider another example falling under the unprotected categories: invasion of privacy. Imagine that Barnes posts Smith's social security number on Twitter. This act does not cause harm so much as it *is* harm. It may be the case that there are no negative consequences as a result of posting this tweet. But that is beside the point: Smith's privacy was undermined in the very act of tweeting. If this is the way in which invasions of privacy harm, then Corlett's own position, according to which utterances falling under the unprotected categories are harmful, compels him to recognize that utterances can sometimes constitute harm, not just cause harm.

Now, Corlett does occasionally use the phrase 'constitute harm' (2018:140). However, he nonetheless seems to conceive of the harmful capacity of utterances in purely causal terms. When he claims that 'hurt feelings caused by hateful expressions do not amount to the causing of genuine harms in the requisite sense', he seems to suppose that the case for the harmfulness of hate speech depends entirely

¹⁹ This example was not chosen at random. Corlett says, "the philosophical assault on the argument in favor of offensiphobia is inspired by John Stuart Mill's view that no political view whatever can legitimately be censured merely because it is offensive" (2018: 121).

on whether hate speech causes harm (2018: 141, our emphasis).²⁰ Hence, it seems to us that Corlett is conflating causal and constitutive harms, while assuming, in a way contrary to the requirements of his own position, that utterances are harmful only if they cause harm.

So, it seems the reason why Corlett believes that proponents of the view that hate speech is categorically harmful assume what they need to argue for is that he mistakenly thinks they need to show that hate speech causes harm.²¹ But many of those who claim that hate speech is harmful defend this claim on the basis that it constitutes harm in ways not covered by the unprotected categories. As it turns out, the list of philosophers cited above claim all believe hate speech is constitutively harmful. For Waldron (2012: 69), hate speech constitutes a genuine harm that is undermining a certain "basic good", namely: 'the assurance of a general commitment to the fundamentals of justice and dignity that a well-ordered society is supposed to furnish to its citizens as part of "the public culture of a democratic society"' (Waldron 2012: 69). Maitra and McGowan (2012: 7) and Langton (2009, 2012) argue that hate speech constitutes harm by subordinating the target (i.e. ranking group members as inferior, depriving group members of powers and capacities, enacting changes to desires, and legitimating discriminatory behavior). And West (2012) argues that hate speech constitutes harm by silencing the target, thereby undermining that target's own right to free speech.

Two features of these arguments for the view that hate speech constitutes harm are noteworthy for our purposes here. First, their proponents are not claiming that the ways in which hate speech is harmful are covered by the unprotected categories.²² They are not usually saying that hate speech causes an unnecessary panic, violence of any kind, or even hurt feelings. Nor are they (only) saying that hate speech constitutes (is) an invasion of privacy or an act of violence.²³ Second, they do not in any way claim that hate speech constitutes harm in virtue of its mere offensiveness. Hence, their arguments cannot be plausibly construed as defenses of offensiphobia. And nor can they be dismissed as question-begging: constitutive harm hate speech theorists are not 'simply repeating, even loudly, that X constitutes hate speech and a harm'; they are offering arguments to this effect—arguments that Corlett fails to engage with accurately or in depth (2018: 142).

²⁰ Consider also the passages we cited earlier: Corlett believes 'it is far from clear that there is a causal link between hateful expressions, on the one hand, and genuine harms, on the other' (2018: 141, our emphasis), and says that 'even if a causal link between hateful expressions and genuine harms could be established... [it would still] include hatefully offensive expressions' that the political left wishes to censure (2018: 141, our emphasis).

²¹ There is some literature—also unacknowledged by Corlett—suggesting that hate speech does, indeed, cause harm (see for example Gelber and McNamara 2016).

²² We think these arguments make a lot more sense of the way in which racist harassment is harmful. Consider Corlett's example of the person who follows a student around campus shouting racist slurs at them (2018: 122). On his account, these racial slurs are harmful only because they violate the right to privacy, and the fact that they are racist seems to be of little or no relevance.

²³ For an argument to the effect that speech might *constitute* an act of violence, see Young (1990, ch. 2).

Our point here is not that any of these arguments are clearly convincing.²⁴ What we have tried to show is that Corlett's apparent belief that such arguments assume what they need to prove might stem from his construal of harm in purely causal terms. It should also be clear that the fact that these arguments might not show that hate speech causes harm does not mean that they assume what they need to prove—they do not *assume* that hate speech *causes* harm; they *argue* that hate speech *constitutes* harm. Perhaps Corlett might still have other reasons for rejecting these arguments, but in order to do this, he needs to argue that they fail to show that hate speech constitutes harm. He cannot simply claim that they are unsuccessful in virtue of not showing that hate speech causes harm. And it is also worth noting that he cannot reject these arguments by claiming that the whole notion of constitutive harm is implausible. This is because, as we have seen, the sorts of utterances he recognizes as harmful, namely, those covered by the unprotected categories, include utterances that can only plausibly be construed as constituting harm (e.g. those which violate privacy).

Let us take stock of what we have shown in this section. We reconstructed Corlett's argument so that it addresses existing expression intolerance rather than the myth of offensiphobia. We saw that it depends on the premise that hate speech is harmful only when it falls under the unprotected categories. We attempted to discern two arguments for this premise from his article: a *reductio* according to which holding that hate speech is harmful would entail that plainly non-harmful utterances are harmful; and an insistence that those who do claim that hate speech is harmful have assumed what they need to prove. We criticized the first argument by pointing out that it relies on a faulty definition of hate speech. We then criticized the second argument by suggesting that it is motivated by Corlett's assumption that utterances are harmful only if they cause harm. This assumption, we saw, is both at odds with existing literature on hate speech and incompatible with Corlett's own position. In light of this, Corlett leaves us with no good reason to think that utterances are harmful only if they fall under the unprotected categories, and that as a result, his argument against the censure of hate speech is unconvincing.

5 Conclusion: the Dangers of the Label 'Offensiphobia'

In this article, we have argued that Corlett's critique of offensiphobia is flawed in at least two ways. First, we have challenged the accuracy of the term 'offensiphobia' as an explanation for the kind of expression intolerance widely present in higher educational institutions and, as a result, shown how Corlett's subsequent critique of this alleged phenomenon attacks a straw man. Second, we have considered how Corlett's remarks on the censure of hate speech, presented in this same article, could be reconstrued as an argument against existing forms of expression intolerance

²⁴ The arguments we have mentioned here are just a sample of the literature on hate speech. If they are found unconvincing, this by no means amounts to a refutation of the entire case for the harmfulness of hate speech.

which call for the censure of hate speech in higher educational settings. We found, however, that this argument is wanting. In this concluding section, we offer a few remarks on how Corlett's critique is especially problematic in light of the difficult situations many universities are currently facing, and how it may even aid the spread of ideologically-distorted beliefs which contribute to this difficult situation.

As we write this article, universities are frequently attacked by governments and influential media, supposedly for the reasons we alluded to in our opening paragraph: universities are now institutions where students and staff neglect the pursuit of truth and good arguments in favor of overindulging each other's sensitivities in a way that diminishes their grasp on reality. While the politicization of university research and teaching is not a new phenomenon, the amount of political muscle which those who share this view have been able to wield certainly seems to have increased dramatically in recent years. In Australia, it was recently announced that the former Minister for Education Simon Birmingham had vetoed \$4.2 million worth of university research grants (incidentally, all from the Humanities), all of which had been approved by the Australian Research Council (Kozoil 2018c). The current federal Minister for Education Dan Tehan is threatening to institute a 'pay for the right to protest' policy (Kozoil 2018a). In Hungary, the recently re-elected government of Victor Orban has banned gender studies programs from universities (Oppenheim 2018). In Brazil, there are reports of police raiding universities, questioning professors, and confiscating material during the lead-up to the election of Jair Bolsonaro as president (Smith 2018).

We expect Corlett is no less troubled by these developments than we are. Yet the propaganda used to justify these actions closely resembles his critique of offensiphobia. At present, conservative politicians and media repeatedly depict university personnel as 'PC snowflakes' who crumble at the first hint of criticism (Quintana 2018). In the wake of the announcement about research funding vetoes in Australia, NSW education minister Rob Stokes launched his own criticism of contemporary university culture, focusing especially on safe spaces and the practice of "no-plat-forming". Apparently, Stokes took it for granted that such practices are forms of offensiphobia, stating:

If, at an intellectual level, you do not have the constitution to cope with hearing views that you find confronting, or even offensive, then you should perhaps reconsider whether a university is the right environment for you to learn in [in Kozoil (2018b)].

Our worry, in sum, is that many of the attacks on higher educational institutions taking place around the world draw political muscle from the idea that many such institutions are hotbeds of offensiphobia, or something like it. Given that offensiphobia is such a blatantly discreditable position (after all, it is a 'phobia', which makes it near certain that no one will ever openly endorse it; even actual homophobes do not usually self-identify as such), this can be used as a convenient justification

for thinly-veiled politically-motivated interventions into academic research and teaching.²⁵

We do not claim that universities are getting everything right. It may well be the case that certain forms of expression intolerance on campuses have been unjustified. Rather, our point is that “offensiphobia” is a red herring. At present, it poses little or no threat to the freedom and integrity of university research and teaching. But labeling universities as offensiphobic may well do so.

Acknowledgements This article benefitted from feedback at the following presentations: “Constructing Hierarchy” Workshop, University of Melbourne (2018); Australian Political Theory and Philosophy Conference, University of Canberra (2019); Exeter Centre for Political Thought seminar series, University of Exeter (2019); Society for Applied Philosophy Annual Conference (2019). We are also grateful to the following people for their comments and advice: Robbie Arrell, Miriam Bankovsky, Thomas Besch, Adrian Currie, Peter Finocchiaro, Moira Gatens, Karen Jones, Matthew Joseph, Michael Longenecker, Sarah Drew Lucas, Timothy Perrine, and Dimitris Vardoulakis. Finally, we are grateful to three anonymous reviewers for their helpful feedback on earlier drafts of this article. This research was partly funded by the Australian Research Council (project ID: DE190100719).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <https://creativecommons.org/licenses/by/4.0/>.

References

- Archard, David. 2014. Insults, Free Speech and Offensiveness. *Journal of Applied Philosophy* 31 (2): 127–141.
- Baker, Edwin C. 2012. Hate Speech. In *The Content and Context of Hate Speech: Rethinking Regulation and Response*, ed. Michael Herz and Peter Molnar, 57–80. New York: Cambridge University Press.
- Bannister, Judith. 2008. It’s Not What You Say but the Way That You Say It: Australian Hate Speech Laws and the Exemption of “Reasonable Expression”. *Florida State University Law Review* 36: 23–40.
- Bray, Mark. 2017. *Antifa: The Anti-Fascist Handbook*. Brooklyn: Melville House.
- Brown, Alexander. 2017a. What is Hate Speech? Part 1: The Myth of Hate. *Law and Philosophy* 36 (4): 419–468.

²⁵ One may ask why we have not mentioned cases of leftist expression intolerance (e.g. no-platforming of Zionist speakers at universities). Our reason is that they seem to be openly motivated by clear political concerns, and not by a desire to curb a culture of ‘offensiphobia’. These political concerns may or may not be plausible (for our part, we are sympathetic to at least some of them). In any case, they do not draw any currency from criticism of offensiphobia in the way that the instances of expression intolerance we have discussed seem to. There may well be contemporary cases where universities are unjustifiably restricting the speech of others, and even that some of these restrictions might be in the name of positions that we are more sympathetic to (i.e. leftist politics). Our point is that offensiphobia is a red herring—it doesn’t help us evaluate these restrictions.

- Brown, Alexander. 2017b. What is Hate Speech? Part 2: Family Resemblances. *Law and Philosophy* 36 (5): 561–613.
- Corlett, J. Angelo. 2018. Offensophobia. *The Journal of Ethics* 22 (2): 113–146.
- Corlett, J. Angelo, and Robert Francescotti. 2002. Foundations of a Theory of Hate Speech. *Wayne Law Review* 48 (3): 1071–1100.
- Davis, Emmalon. 2018. On Epistemic Appropriation. *Ethics* 128: 702–727.
- Dworkin, Ronald. 2009. Foreword. In *Extreme Speech and Democracy*, ed. Ivan Hare and James Weinstein, 5–9. Oxford: Oxford University Press.
- Erjavec, Karmen, and Melita Poler Kovičič. 2012. "You Don't Understand, This is a New War!" Analysis of Hate Speech in News Web Sites Comments. *Mass Communication and Society* 15 (6): 899–920.
- Feinberg, Joel. 1980. *Rights, Justice, and the Bounds of Liberty*. Princeton: Princeton University Press.
- Feinberg, Joel. 2004. Limits to the Free Expression of Opinion. In *Philosophy of Law*, 7th ed, ed. Joel Feinberg and Jules Coleman, 379–395. Belmont: Wadsworth.
- Finlayson, Lorna. 2012. 'Free Speech' Noise and the Silence of David Willets. Posted 17 February 2012. <https://donsspeakout.wordpress.com/2012/02/17/lorna-finlayson-free-speech-noise-and-the-silence-of-david-willetts/>. Accessed 28 Nov 2018.
- Fricker, Miranda. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.
- Gelber, Katharine. 2019. Differentiating Hate Speech: A Systemic Discrimination Approach. *Critical Review of International Social and Political Philosophy*. <https://doi.org/10.1080/13698230.2019.1576006>. (online first).
- Gelber, Katharine, and Luke McNamara. 2016. Evidencing the Harms of Hate Speech. *Social Identities: Journal for the study of Race, Nation and Culture* 22 (3): 324–341.
- George, Cherian. 2014. Journalism and the Politics of Hate: Charting Ethical Responses to Religious Intolerance. *Journal of Mass Media Ethics* 29 (2): 74–90.
- Greene, Jamal. 2012. Hate Speech and the Demos. In *The Content and Context of Hate Speech: Rethinking Regulation and Response*, ed. Michael Herz and Peter Molnar, 92–115. New York: Cambridge University Press.
- Haraszti, Miklos. 2012. Foreword: Hate Speech and the Coming Death of the International Standard Before it was Born (Complaints of a Watchdog). In *The Content and Context of Hate Speech: Rethinking Regulation and Response*, ed. Michael Herz and Peter Molnar, 13–18. New York: Cambridge University Press.
- Harvey, Andy. 2012. Regulating Homophobic Hate Speech: Back to Basics about Language and Politics? *Sexualities* 15 (2): 191–206.
- Herszenhorn, David M. 2006. Graduates at New School Heckle Speech by McCain. The New York Times, May 20 2006. <https://www.nytimes.com/2006/05/20/nyregion/20mccain.html>. Accessed 28 Nov 2018.
- Kozoil, Michael. 2018a. You protest, you pay: Education Minister's bid to bolster free speech at universities. *The Sydney Morning Herald*, September 22. <https://www.smh.com.au/politics/federal/you-protest-you-pay-education-minister-s-bid-to-bolster-free-speech-at-universities-20180921-p5057h.html>. Accessed 28 Nov 2018.
- Kozoil, Michael. 2018b. Education Minister: Foreign students a threat to academic independence. *The Sydney Morning Herald*, October 22. <https://www.smh.com.au/politics/federal/education-minister-foreign-students-a-threat-to-academic-independence-20181021-p50azp.html>. Accessed 28 Nov 28, 2018.
- Kozoil, Michael. 2018c. Former education minister vetoed \$4.2 million in recommended university research grants. *The Sydney Morning Herald*, October 26. <https://www.smh.com.au/politics/federal/former-education-minister-vetoed-4-2-million-in-recommended-university-research-grants-20181026-p50c3a.html>. Accessed 28 Nov 2018.
- Langton, Rae. 2009. *Sexual Solipsism*. Oxford: Oxford University Press.
- Langton, Rae. 2012. Beyond Belief: Pragmatics in Hate Speech and Pornography. In *Speech and Harm: Controversies over Free Speech*, ed. Ishani Maitra and Mary Kate McGowan, 72–93. Oxford: Oxford University Press.
- Maitra, Ishani, and Mary Kate McGowan. 2012. Introduction and Overview. In *Speech and Harm: Controversies over Free Speech*, ed. Ishani Maitra and Mary Kate McGowan, 1–23. Oxford: Oxford University Press.
- Matsuda, Mari. 1989. Public Response to Racist Speech: Considering the Victim's Story. *Michigan Law Review* 87 (8): 2320–2381.

- McGowan, Mary Kate. 2018. On Political Correctness, Microaggressions, and Silencing in the Academy. In *Jennifer Lackey*, ed. Academic Freedom, 135–146. Oxford: Oxford University Press.
- Mill, John Stewart. 1978 [1859]. In *On Liberty*, ed. Elizabeth Rapaport. Cambridge: Hackett Publishing Company, Inc.
- Oppenheim, Maya. 2018. Hungarian Prime Minister Viktor Orban Bans Gender Studies Programmes. *The Independent*, October 24. <https://www.independent.co.uk/news/world/europe/hungary-bans-gender-studies-programmes-viktor-orban-central-european-university-budapest-a8599796.html>. Accessed 28 Nov 2018.
- Quintana, Chris. 2018. Colleges Are Creating ‘a Generation of Sanctimonious, Sensitive, Supercilious Snowflakes,’ Sessions Says. *The Chronicle of Higher Education*, July 24. <https://www.chronicle.com/article/Colleges-Are-Creating-a/243997>. Accessed 28 Nov 2018.
- Rabban, David M. 1994. The IWW Free Speech Fights and Popular Conceptions of Free Expression Before World War I. *Virginia Law Review* 80 (5): 1055–1158.
- Reger, Jo. 2015. The Story of a Slut Walk: Sexuality, Race, and Generational Divisions in Contemporary Feminist Activism. *Journal of Contemporary Ethnography* 44 (1): 84–112.
- Richardson-Self, Louise. 2018 (forthcoming Dec). Offending White Men: Racial Vilification, Misrecognition and Epistemic Injustice. *Feminist Philosophy Quarterly*. https://www.academia.edu/36547033/Offending_White_Men_Racial_Vilification_Misrecognition_and_Epistemic_Injustice. Accessed 28 Nov 2018.
- Schauer, Frederick. 2012. Social Epistemology, Holocaust Denial, and the Post-Millian Calculus. In *The Content and Context of Hate Speech: Rethinking Regulation and Response*, ed. Michael Herz and Peter Molnar, 129–143. New York: Cambridge University Press.
- Schulzke, Marcus. 2016. The Social Benefits of Protecting Hate Speech and Exposing Sources of Prejudice. *Res Publica* 22: 225–242.
- Simpson, Robert, and Amia Srinivasan. 2018. No Platforming. In *Jennifer Lackey*, ed. Academic Freedom, 186–206. Oxford: Oxford University Press.
- Smith, Amy Erica. 2018. Brazilian Media Report that Police are Entering University Classrooms to Interrogate Professors. *Vox*, November 1. <https://www.vox.com/mischiefs-of-faction/2018/10/26/18029696/brazilian-police-interrogate-professors>. Accessed 28 Nov 2018.
- Waldron, Jeremy. 2012. *The Harm in Hate Speech*. Cambridge: Harvard University Press.
- West, Caroline. 2012. Words that Silence? Freedom of Expression and Racist Hate Speech. In *Speech and Harm: Controversies over Free Speech*, ed. Ishani Maitra and Mary Kate McGowan, 222–248. Oxford: Oxford University Press.
- Young, Iris Marion. 1990. *Justice and the Politics of Difference*. Princeton: Princeton University Press.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.