Theoria, 2025, 40(1), 30-46
https://doi.org/10.1387/theoria.26849

THEORIA
ISSN 0495-4548 – eISSN 2171-679X

The harms of non-derogatory uses of slurs and the Potential Normalization Argument

(Los daños de los usos no derogatorios de los slurs *y el Argumento de la Normalización Potencial)*

Dan Zeman*
University of Porto
https://orcid.org/0000-0001-5620-0042

Keywords

Slurs
Derogatory and non-derogatory uses
Harm
Normalization

ABSTRACT: There is no doubt that slurs harm. They do so by denigrating their targets, by putting them down, by marginalizing them. This is why in many legislations around the world, the use of slurs has been banned or penalized. But should *all* uses of slurs be banned? Many uses of slurs seem to be non-derogatory and to have beneficial effects. However, such uses are double-faceted: as both armchair reflection and experimental studies have shown, they are able to produce harm as well. In this paper, I approach the broad question of whether all non-derogatory uses of slurs should be banned. I first present the main uses of slurs that have been considered to be non-derogatory and recent reactions to those. The upshot of this survey is that uses of slurs that have been considered non-derogatory do, in fact, produce harm. I also flag what various authors have recommended in relation to the issue of banning such uses. Against this background, I engage with a recent view put forward by Alba Moreno Zurita and Eduardo Pérez-Navarro, who urge extreme caution with respect to any uses of slurs, due to their potential to normalize derogation. After presenting their view and their main argument, I raise an objection related to their treatment of neutral uses of slurs. I end with pointing out that, while their endeavour has merit in that it pushes the discussion further, it raises certain issues —of both an empirical and a normative nature— that need to be addressed.

Palabras clave

Slurs
Usos derogatorios y no derogatorios
Daño
Normalización

RESUMEN: No cabe duda de que los *slurs* hacen daño. Lo hacen denigrando a sus destinatarios, menospreciándolos, marginándolos. Por eso, el uso de *slurs* ha sido prohibido o penalizado en muchas legislaciones de todo el mundo. Pero ¿deberían prohibirse todos los *slurs*? Muchos de sus usos parecen no ser derogatorios y tener efectos beneficiosos. Sin embargo, tales usos tienen dos caras: como han demostrado tanto la reflexión de sillón como los estudios experimentales, también tienen la capacidad de producir daño. En este artículo me planteo la cuestión general de si deberían prohibirse todos los usos no derogatorios de los *slurs*. Primero, presento los principales usos de los *slurs* que se han considerado no derogatorios y las principales reacciones recientes a los mismos. El resultado de este estudio es que los usos de los *slurs* que se han considerado no derogatorios, de hecho, producen daño. También señalo lo que varios autores han recomendado en relación con la cuestión de prohibir tales usos. En concreto, me ocupo de una posición reciente presentada por Alba Moreno Zurita y Eduardo Pérez Navarro, quienes instan a extremar la precaución respecto a cualquier uso de los *slurs*, debido a su potencial para normalizar la derogación. Tras presentar su punto de vista y su argumento principal, planteo una objeción relacionada con el tratamiento que dan a los usos neutrales de los *slurs*. Concluyo señalando que, si bien su esfuerzo es meritorio en la medida en que hace avanzar el debate, plantea ciertas cuestiones —de naturaleza tanto empírica como normativa— que es necesario abordar.

* ***Correspondence to:*** Dan Zeman. Institute of Philosophy, Faculty of Letters, University of Porto, Via Panorâmica, s/n, 4150-564, Porto, Portugal – zeman@letras.up.pt – https://orcid.org/0000-0001-5620-0042

How to cite: Zeman, Dan (2025). «The harms of non-derogatory uses of slurs and the Potential Normalization Argument»; *Theoria. An International Journal for Theory, History and Foundations of Science,* 40(1), 30-46. (https://doi.org/10.1387/theoria.26849).

Received: 07 August, 2024; Final version: 24 April, 2025.

ISSN 0495-4548 - eISSN 2171-679X /

There is no doubt that slurs harm. They do so by denigrating their targets, by putting them down, by marginalizing them. Worse still, slurs can be an aid to dehumanization and sometimes they can play a crucial role in bringing about violent outcomes such as genocide. Additionally, experimental studies in psychology have conclusively shown that slurs affect their targets and lead to a decrease in their well-being, by influencing their self-worth and performances, and increasing stereotype- and identity-threat. This is why in many legislations around the world, slurs (or at least their derogatory uses) have been banned or penalized.

But should *all* uses[1] of slurs be banned? First, at least on the face of it, it seems that using slurs in some contexts can be beneficial. For example, when a court is investigating, say, a case involving verbal aggression, it might be crucial for the jury to know exactly what has been said in order to reach the right verdict. Similarly, when journalists report a story involving the proffering of slurs, it might be important to reproduce the same words to make the information as exhaustive as possible in order to arrive at a maximally detailed description of the facts. Finally, in research and teaching, those involved in these activities often use examples to introduce their topic of study (slurs, pejoratives, hate speech, etc.), to get a certain point across, or to support an argument. In all these cases, slurs are being used (although in quotes, or by using other detachment devices). Is it reasonable to ban such uses?

Second, and perhaps more substantially, many discriminated communities have used slurs to overcome precisely the discrimination they have been historically subjected to, appropriating or reclaiming the word that has been used as a weapon against them. Changing a slur by endowing it with positive meaning instead of a negative one has been a powerful tool in this process. Furthermore, as several authors have argued, slurs can also be employed neutrally —as when a slur is used to refer to someone's friends, or as a mere tag to self-identify as belonging to a certain group (an ethnic one, for example). Should *these* uses be banned?

Arguably, the uses of slurs listed above are non-derogatory —or at least have been considered so. While the case for banning derogatory uses of slurs across the board is relatively easy to make (due to their obvious pernicious effects, or the harm they produce[2]), the same cannot be said about non-derogatory uses. The issue is complicated, because non-derogatory uses of slurs seem to be double-faceted: on the one hand, as we just saw, such uses of slurs can be beneficial, and depending on the type of use, they can play a more or less important role in a community; on the other hand, it has been argued (as we will see shortly) that some of what are taken to be non-derogatory uses of slurs can have pernicious effects, too. So, the question becomes, in what contexts and under which circumstances should purported non-derogatory uses of slurs be banned? This is the broad issue I address in this paper.

To approach it, I will first expand on what were thought to be, at different points in time and by various authors, non-derogatory (and thus non-harmful) uses of slurs, together with more recent reactions to such uses (section 1). What

---

1 I'm employing here the term "use" to refer to any occurrence of a slur (either oral or in writing), regardless of the linguistic embedding or non-linguistic context. One consequence of this usage is that the familiar "use- mention" distinction that philosophers of language often warn about collapses; for me, mentioning a slur is one type of use of that slur, regardless of the exact mechanism by which mentioning is communicated (quotes, air quotes, tone, implicit understanding, etc.). It is perhaps worth noting that, despite its treatment being a contentious issue, this take on the use-mention distinction is consistent with certain theoretical approaches to quotation (e.g., Frege's *Identity Theory*; see García-Carpintero, 2004, p. 678). I embrace this broad sense of "use" because although I will talk about both cases of mentioning slurs and of slurs being used in non-derogatory ways without being mentioned, I'm interested in the question whether *all* non-derogatory uses of slurs should be banned.

2 I'm operating in this paper with a very loose notion of harm, one that comprises any (systematic, targeted and interpersonal) negative effect a slur can have on a member of the target group. This opens up at least a couple of important issues: whether any such harm should be taken into consideration when banning a certain word is discussed and whether the harm produced can be considered to be the direct result of the semantic profile of a slur or of other, perhaps pragmatic or simply sociological, factors. I take here the stance (without arguing for it) of allowing any (systematic, targeted and interpersonal) harm to be at least a prima facie good reason to consider banning a certain word. As for the second issue, I remain neutral on the issue of whether the relevant harm is necessarily a result of the semantic profile of a slur or not, as I don't think anything I will say in what follows trades on taking one view or another. All I assume is that there is a connection between harm and the derogatory character of a slur, regardless of how this is spelled out theoretically.

emerges from considering these reactions is that non-derogatory uses of slurs can also have pernicious effects and thus cause harm. After flagging what the respective authors recommend vis-à-vis various uses of slurs previously considered to be non-derogatory, I move on to the main focus of this paper: a recent, more general, argument found in the work of Alba Moreno Zurita and Eduardo Pérez-Navarro who argue that *any* use of a slur is potentially harmful, and thus urge caution when using slurs. I present their main argument (which I reconstruct and dub "the Potential Normalization Argument") and other important claims they make in section 2, while in section 3 I bring forth an objection to their view. I take the argument to be an important one to discuss, regardless of whether the objection is successful. Thus, while this paper is quite limited in scope, I take it that engaging with their view pushes the discussion in new, interesting directions.

1. The uses of slurs and their harms

While it is widely agreed that the main function of slurs is "to derogate or dehumanize (...), to signal that their targets are unworthy of equal standing or full respect as persons" (Jeshion, 2013, p. 232), they have many uses, not all of them derogatory. As a broad two-partite categorization, one can group uses of slurs in *derogatory* (also called "weapon uses" by Jeshion, 2013) and *non-derogatory* ("non-weapon uses"). In what follows, I will take each broad category in turn, underlying some of the ways they have been considered to produce harm. In connection to non-derogatory uses, I will showcase several such presumed uses, and flag what various authors reacting to them have recommended vis-à-vis their banning.

1.1. Derogatory uses and their harms

Derogatory uses are the closest to fulfilling the main function of slurs spelled out above, and they are illustrated by sentences like "John is a boche."[3], "You are a boche.", "You, boche!" and so on.[4] The pernicious effect of derogatory uses of slurs is undeniable, and it has been amply documented, with the types of harms they produce varying from severe forms to less so.

Thus, as work on dehumanization shows (see chiefly Livingstone Smith, 2011, 2020, 2022), slurs have been one of the tools by which people dehumanize those they deem inferior. Dehumanizing one's (perceived) enemies makes it easier to isolate, discriminate, and marginalize them, as not being on the same level on the Great Chain of Being. Further, slurs are an effective way to help bringing about violent outcomes when parts of societies clash: as the widely cited work by Tyrrell (2012) concerning the Rwandan genocide shows, the use of slurs and other words such as "cockroaches" and "snakes" has been a crucial aid to starting and justifying the massacre of the Tutsi and of other minorities by the Hutu. Less dramatically, but not less importantly, experimental studies in psychology have shown that being targets of slurs leads to internalization of negative attitudes and associations (Carnaghi & Maass, 2008), to feelings of alienation due to exclusion and to a "us vs. them" frame of mind, which in turn lead to a decrease in self-assuredness and self-esteem (Carnaghi & Maass, 2008), and to stereotype threat, which leads to worse performances by the members of the target group due to worries about conforming to the stereotype (Steele *et al.*, 2002).[5] As can be seen from this list, while not all uses of slurs end up having dramatic effects such as dehumanization or genocide, even the most mundane derogatory use of a slur can affect its target in significant ways.

---

[3] Whether to use slurs (in quotation marks) in academic papers is one of the issues this paper deals with. As can be seen, here I take the widespread position that mentioning some slurs is important.

[4] There are nuances about the different ways such constructions derogate: for example, *calling* someone a boche instead of merely *asserting* that they are a boche might have a slightly different derogatory profile. Interesting as they are, I will leave such issues aside.

[5] I'm following here the very useful review of the experimental literature in (Herbert, 2017).

It is generally contended that at least part of the harm produced by slurs comes from normalizing certain ways of thinking and talking about the target group, which makes it easy to act on the beliefs that possibly lead to the outcomes described in the previous paragraph. In this, slurs have been thought to be intimately connected with pernicious ideologies, with the relation between them being one of *mutual reinforcement*: the derogatory use of slurs is made possible by the ideologies that consider members of the target group to be inferior as humans, while repeated use further strengthen those ideologies, perpetuating negative stereotypes about the members of the target group and essentializing them.[6] It is also important to note that slurs do so often implicitly, escaping people's conscious efforts and relying instead on their biases (Herbert, 2017).

1.2. Non-derogatory uses and their harms

Non-derogatory uses of slurs come in many forms.[7] Below I present several types of uses of slurs that have been taken to be non-derogatory, together with more recent reactions to them. Subsection 1.3. discusses the upshot for the issue tackled in this paper.

1.2.1. Negation

Denying that a slur applies to someone by means of sentences like "Hans is not a boche." or "There are no boches; there are only Germans." has been considered a non-derogatory use of the slur. Thus, such sentences have been taken as "disowning the [derogatory] attitude" (Blackburn, 1984, p. 148), to "convey that ['boche'] is not something one calls anyone" (Hornsby, 2001, p. 129) or to be a way of "repudiating the use of the term" (Dummett, 2007, p. 527). Other authors take the negation in such sentences to be metalinguistic (e.g., Hom, 2008).

However, despite assurance from these philosophers, it has been remarked that denying that a slur applies to someone by using negation (as in the sentence "Hans is not a boche.") is not the most efficient way to go when one intends to defuse the effect of a slur, as its presence, even within negation, manages to offend (the speaker is using the slur, even if their intentions are to defuse it). Anderson & Lepore (2013) thus claim that "a denial of ['Hans is a boche.'] is no less inflammatory than ['Hans is a boche.'] itself and don't recommend this strategy when attempting to reject a derogatory, weapon-use of a slur" (p. 28).

1.2.2. Speech and belief reports

Disquoting a sentence in an indirect speech or a belief report is usually a correct way to convey what the reported person said or believed. Thus, reporting utterances containing a slur has been considered non-derogatory, the embedding of the slur under the speech or attitude verb insulating its derogatory character (in the sense that it is not the speaker who derogates, but the person whose speech or belief is reported). Schlenker (2003) has provided some examples purporting to show the insulating effect of such reports, the most well-known being "I am not prejudiced against Caucasians. But John, who is, thinks/claims that you are the worst honky he knows" (Schlenker, 2003, p. 43).

As with negation, it has been recently contended that what has been claimed to be an insulating effect of speech and attitude verbs is, in fact, a defective environment (with Schlenker's example being heavily contested). What many authors

---

6 For concrete proposals about how the two relate, see Hom (2008), Kukla (2018), Popa-Wyatt & Wyatt (2018), Davis & McCready (2020), Swanson (forthcoming) —among many others.

7 For a list, partially repeated here, see Cepollaro & Zeman (2020). That list also contained fictional uses of slurs (both as in real world slurs used in fiction —e.g., the n-word used in movies, and as slurs created for fictional characters— e.g., the slur "toaster" used for replicants in the series *Battlestar Galactica*). Slurs used in humour or for comedic effect can also be counted among non-derogatory uses. Although they are important and raise a variety of issues, I will ignore them in what follows. See, however, Anderson (2015) for an introductory discussion of the topic.

agree about is that, in reporting a use of a slur, the speaker cannot simply claim innocence, since the slur is being used in the report as well and derogation is traced to the speaker. Additionally, there is ample recent experimental evidence that reported slurs, while less derogatory than non-embedded ones, are nevertheless perceived as offensive (Cepollaro *et al.*, 2019; Cepollaro *et al.*, 2024). In fact, it is nowadays agreed that slurs "scope-out" from any kind of linguistic environments —including conditionals like "If Hans is a boche, we shouldn't invite him for dinner." This characteristic of slurs has been dubbed "hyper-projectivity" or "non displaceability" (Potts, 2007; Hom, 2010; Croom, 2011; Cepollaro, 2020; etc.) and refers to the fact that the derogatory character of slurs doesn't disappear even when they are embedded in linguistic constructions. As the examples discussed in these paragraphs show, despite slurs appearing embedded under negation, indirect speech and belief reports, and conditionals, they retain their derogatory character —just as they do when used unembedded.

1.2.3. Quotation

Many uses of slurs are didactical or quotational —such as those already mentioned at the beginning of this paper. Quotation, the widespread device by which one insulates the meaning of words to focus on the words themselves, and which is the predominant tool in court, newspapers and classrooms, has once been thought to provide a safe haven for slurs. Thus, Hom takes "explicit quotation in the courtroom" to be among acceptable uses of slurs (2008, p. 427), Williamson believes that "mere quotation marks (...) isolate us from [a slur's] derogatory implications" (2009, p. 139), while Hornsby claims that "quotation has some sealing off effect" (2001, p. 130).

Despite widespread past consensus that quotation isolates the derogatory character of slurs, it has, again, been noted that quoted slurs can derogate as well. Famously, and in opposition to (the then) orthodoxy, Anderson & Lepore note that the occurrence of a slur in a sentence like "'Boche' is a term for Germans" "can easily cause alarm and offence" (2013, p. 37) and counter the more lenient claims of Hom, Williamson and Hornsby by noting that, even in such contexts, there is a preference for using more acceptable formulations (e.g., "the n- word") instead of the slur (in this case, the n-word). Thus, in relation to such uses, they propose the following strategy: "As a safeguard against such inurnment, we strongly urge you always to ask yourself how a targeted member, perhaps accidentally overhearing you, would react to your usage. You'll find, as we have, that much of what seems suitable is definitely not" (Anderson & Lepore, 2013, p. 31).

This noted tendency of slurs' derogatory aspect to "scope-out" of quotation and produce harm is supported by empirical research. First, in the same study mentioned above in relation to speech reports (Cepollaro et al, 2024), the authors have found that slurs reported both in direct speech ("Z said that 'Y is an S'") and in mixed quotation ("Z said that Y is 'an S'"), while less derogatory than non-embedded ones, are still perceived as offensive. Second, in some of the experimental studies surveyed by Herbert (2017), participants were subjected to single printed slurs (Carnaghi & Maass, 2008) or received subliminal exposure of slurs (Fasoli *et al.*, 2016), and a similar effect has been found to participants being subjected to uses of slurs in a context. Herbert takes this to show that "both using and mentioning the slur can prime audience members in this harmful way" and that "[m]erely mentioning a slur can act as a powerful priming mechanism of pernicious implicit associations" (2017, p. 144). While this result has been achieved in testing outgroup members, it is reasonable to claim that it also holds for members of the target group. To put it succinctly, the experimental evidence she samples shows that there is not much difference between uses of slurs in context and mere occurrences of slurs in terms of the harm produced; in her view, this "raises important ethical questions for how we ought to talk about slurs" (Herbert, 2017, p. 147).

1.2.4. Appropriation

A great number of slurs have been appropriated (reappropriated, reclaimed —the terminology varies; I will use these terms here interchangeably) or are in the process of being so. For example, "queer" is often cited as a successful instance of an appropriated slur (as the titles of journals like *Queer Studies* or of TV shows like *Queer Eye* attest), as well as the n-word, when used within certain communities in the U.S.A. Appropriation itself has many facets, and scholars have distinguished at

least two types: i) appropriation within political activism; ii) appropriation as signalling or fostering camaraderie, with many other uses being attested too (Naylor, 1986; Kennedy, 2003; Jeshion, 2013, 2020; Bianchi, 2014; etc.).

What about these uses? While generally the use of slurs within the process of appropriation has been looked upon favourably, there have been voices pointing out that reclamation projects have their own risks. First, there are various reactions to appropriating specific slurs, as the history of the n-word, of "queer" or of "slut" attests. Thus, opposition to the appropriation of the n-word has been present in the African-American cultural space from the beginning (see, among many others, (Asim, 2007) and (Rahman, 2012) for discussion); Brontsema (2004) details the history of the appropriation of "queer", contextualizing the inner struggles within the target community in relation to what is nowadays considered the reclaimed slur par excellence; "slut", on the other hand, is considered an example of a failed attempt at reclaiming a slur, facing strong opposition despite gathering wide support through the famous "slut walks" (see, e.g., Kleinman *et al.*, 2009 or Black Women's Blueprint, 2016).[8] In all these cases, using the slur in the process of appropriation was considered by its opponents as producing harm by reinforcing the dominant, oppressive ideology that the appropriation was supposed to fight against.

On a more general level, Herbert (2015) argues that reclamation is intrinsically a "precarious project". Operating within a speech act framework, Herbert investigates the pragmatic structure of reclamation projects and argues that they face hazards that are distinct from other forms of protest. She points out that reclamation can fail, which in turn leads to harm by reinforcing the very mechanisms of oppression against which the process has started in the first place. Thus, she remarks: "This is the precarious structure of reclamation projects: when successful, reclamation is the subversion of powerful mechanisms of oppression, but when unsuccessful, the act has the ironic force of constituting mechanisms of oppression." (Herbert, 2015, p. 132). However, even though she is wary of reclamation projects, Herbert doesn't argue for banning appropriated or reclaimed uses: "Reclamation is intrinsically hazardous. Yet it can also be worthwhile: even when reclamation projects fail as reclamation, they still may accomplish good. (...) While it will always be precarious, it may on occasion be well worth the risks" (2015, p. 137).[9]

It is interesting to note that appropriated uses of slurs have been seen favourably even by proponents of views that are considered more radical when it comes to cautioning against non-derogatory uses of slurs. For example, as we have seen, under Anderson & Lepore's prohibitionist view, slurs —even in embeddings or quoted— should be banned. However, when it comes to appropriation, they write that "in cases of appropriation, a target group member can opt to use a slur without violating its prohibition because his membership provides a defeasible escape clause" (Anderson & Lepore, 2003, 42). Another class of (radical) views —such as Rappaport's (2020) or Stojnić & Lepore's (2022)— has in common the idea that producing harm is tied to the phonological realization of slurs rather than having to do with their content, either semantic or pragmatic. Yet, when addressing appropriation, Rappaport writes that "in cases of appropriation, it's arguable that a slur can retain its toxicity, despite shifting its conventional use, and such cases may certainly be justifiable" (2020, p. 199), while Stojnić & Lepore (2022) claim that "changes in association over time could also lead to the articulation becoming tied to fully neutral or positive associations, leading to an effective lifting of the ban" (p. 753).

1.2.5. Neutral uses

A different group of non-derogatory uses of slurs consists in them being used neutrally. Anderson (2018), relying on work by Smitherman (2006), points out that within certain communities of African Americans in the U.S.A. the n-word is used to mean "friend", "buddy", "man"; he dubs this type of use *referential*. In Zeman (2022), I have proposed

---

[8] See, however, (Gaucher *et al.*, 2015) for an experimental study showing that the appropriation of "slut" can have positive effects, such as lowering the likelihood of women endorsing rape myths and empowerment leading to positive impact in collective action.

[9] This stance is, of course, consistent with the experimentally proven fact that, when successful, appropriation of a slur has positive effects on the member of the target group. See, among others, (Galinsky *et al.*, 2013).

to interpret a certain use of the ethnic slur "țigan" from Romanian and other Eastern European languages (roughly translated in English as "gipsy") by members of the Roma community as a neutral use; since the aim of such a use is for the members of the community to identify as belonging to a certain ethnic group, I dubbed this use *identificatory*. The main difference between neutral and appropriated uses is that, while the latter aim to put a positive spin on the negative meaning of a slur, the former are used without such an aim.

Neutral uses have not been focused on so far in the literature, so it is difficult to discern what the prevalent attitude towards them would be. While one could extrapolate from what one author or another thinks about non-derogatory uses of slurs in general or about certain types of such uses (e.g., appropriated uses), the results of doing so don't seem promising. However, I take the failure to address such uses as a drawback, since —as I will show— they raise important issues vis-à-vis our stance on banning non-derogatory uses of slurs, both for the authors surveyed so far and for the two who constitute the main focus of this paper.[10]

1.3. A preliminary conclusion

Before moving on to the main argument I want to tackle in this paper, let's take stock. The preceding subsection shows that the harm produced by slurs extends well beyond their derogatory uses. The picture that emerges is that many of the uses of slurs that have been thought to be non-derogatory are not considered as such anymore, while the evidence that they, in fact, produce harm, could form the basis of an argument for banning them.

However, despite this blanket agreement, things get complicated when one considers particular non-derogatory uses and the surveyed authors' attitudes towards them —as even a cursory look at the showcased attitudes and recommendations shows. What seems to transpire is that the authors cited above have slightly different attitudes towards different uses that have been considered non-derogatory. For example, while Anderson & Lepore, Rappaport and Stojnić & Lepore advise against using slurs in quotation, Herbert only urges caution in using them in this way (at the same time offering comprehensive advice on how to mitigate their negative effects; see Herbert, 2017). Appropriated uses have generally seen favourably, even though some authors are keen on pointing out the risks associated with this endeavour (e.g., Herbert, 2015). Thus, even though the authors mentioned converge in the general attitude of cautioning against unrestricted proffering of non-derogatory uses of slurs, there is still some disagreement about the exact nature of the attitude recommended in relation to *particular* types of such uses and under which conditions they should be avoided. This situation seems to preclude both drawing a unitary conclusion vis-à-vis non-derogatory uses of slurs from the ensemble of work investigated and claiming that a general ban on such uses is mandated.

There are other issues to consider as well. For example, if a less lenient attitude is recommended towards one type of non-derogatory use of slurs (say, slurs used in quotation or in reports), but a more permissive one when it comes to another type (say, appropriated uses), then one question that arises is what happens when these two types of uses are combined —for example, when quoting and reporting slurs are part of a process of appropriation. To put it differently, the issue here is that it is unclear what is the "resulting" attitude one should have in such (and similar) cases, and without settling this it is not clear whether the type of use(s) in question should be banned or not. One can claim that, in *this* particular combination of types of uses, the fact that slurs are used, in whatever ways, within the larger umbrella of a process of appropriation makes them free for use. But this is, first, a substantial claim (that appropriation trumps any

[10] Other non-derogatory uses of slurs found in literature are what Hom has called "non-derogatory non- appropriated" uses, illustrated by examples like "Institutions that treat Chinese as chinks are morally depraved." (2008, p. 423), aimed at protesting against or correcting institutional wrongs. There are also metaphorical non-derogatory uses of slurs, as the title of a famous song by John Lennon and Yoko Ono illustrates; the aim of using the n-word here has been not to derogate a racial group (African Americans), but to draw attention to the similar condition of another (women). Jorgensen Bolinger (2020) discusses uses of slurs by competent speakers of a language that are not convinced that the words they use are derogatory. Although interesting, I leave such uses aside, as they have not been discussed in the literature in connection to the issue I'm concerned with here.

other type of use) and, second, something that none of the authors mentioned has explicitly stated. Whatever the answer to particular combinations of types of uses might be, it is not trivial.

Additionally, many of the authors mentioned above have not dealt with neutral uses of slurs. As I stressed above, the difference between neutral and appropriated uses is that in the latter, but not in the former, the resemantization of a slur presupposes a change from negative to positive, which doesn't need to happen in neutral uses. Treating these two types of uses in a similar way might be prima facie reasonable, but (as I will claim in the next section) neutral uses might have certain characteristics that appropriated uses do not have, and thus giving them a common treatment might lead to an incomplete or otherwise inadequate theory. In any case, as with the combination of different types of non-derogatory uses of slurs, this issue has largely remained unaddressed in the literature, and the solution might not be a straightforward one.

Now, many of the authors I engaged with above have, at various times, motivated apermissive attitude towards a type of non-derogatory use or another by claiming that such uses are "worthwhile" and "worth the risk" (Herbert), "justified" or "justifiable" (Rappaport), or governed by a "defeasible escape clause" (Anderson & Lepore). The leading idea behind a permissive attitude thus seems to be that there is something to be gained from a particular use, even if some harm is produced; in most cases, the envisaged advantage is practical or social (in a slogan: a use of a slur is kosher if it is in the service of a good cause). This is certainly a reasonable approach. However, an encompassing theory of what this amounts to, both in general and with respect to particular types of non-derogatory uses, is not easy to find in the literature. What seems to be needed, thus, is a *general* argument of why and when non- derogatory uses of slurs are permissible. In the next section, I turn to a recent theory that aims to offer such an argument, based on the idea of the benefits of using slurs without ignoring their harmful potential, and one that considers their neutral uses as well.

2. The Potential Normalization Argument

In recent work, Alba Moreno Zurita and Eduardo Pérez-Navarro (henceforth "MZ&PN") have urged for caution in relation to any way of using slurs. Their main claim is that "slurs are always potentially harmful, even if some of their occurrences are non-derogatory" (Moreno & Pérez-Navarro, 2021, p. 708). They take this observation to support the stance of not using slurs, even in non-derogatory contexts. Interestingly, and highly relevant for the issue I consider, their position is not a *silentist* one: they do not intend to ban slurs tout court. Instead, what they suggest is that every use of a slur comes with a moral cost, and that it is the speakers' decision (and, hence, responsibility) to determine whether the benefits of using a slur outweigh the cost. What their work adds to the discussion is, first, that they consider non-derogatory uses of slurs that have not been considered in the literature (i.e., *neutral* uses of slurs), and that their argument for being cautious in using slurs non-derogatorily is different from the extant ones in the literature in that it is more general. In this section, I expand on their claims and put forward what I take their main argument to be, which (for reasons that will become clear shortly) I dub the "Potential Normalization Argument".

Thus, MZ&PN start with distinguishing two types of contexts in which slurs can be used: *controlled* contexts and *uncontrolled* contexts. The latter are "contexts in which we do not have enough knowledge of our audience to predict what the uptake of the utterance will be" (Moreno & Pérez-Navarro, 2021, p. 710), due to the fact that most factors that impact the effect of an utterance are beyond our control; the former are contexts in which "we can predict with reasonable accuracy what the consequences of a given utterance will be" (Moreno & Pérez-Navarro, 2021, p. 715), due to our familiarity with the audience. They think that this distinction overlaps with that between derogatory and non-derogatory uses of slurs; that is, while slurs always[11] derogate when used in uncontrolled con-

[11] MZ&PN address cases of uncontrolled contexts in which all the members of the audience take the use of a slur to be non-derogatory by appealing to the idea (which they borrow from Lasersohn, 2007) that "hearers have *every reason* to attribute to [the speaker] a negative attitude toward the target group" (Moreno & Pérez-Navarro, 2021, p. 718). I skip the discussion of what exactly that amounts to, as it doesn't bear on the issues I want to raise.

texts, they can be used non-derogatorily in controlled ones (the claim is not that they always are used like this in such contexts —the use of a slur in a controlled context formed only by members of a hate group is still derogatory). As examples of controlled contexts, MZ&PN provide two scenarios, one involving a pedagogic use of a slur, the other involving an ironic use. Here is how they describe the two situations, as well as their comments on them (labels mine):

> THE PEDAGOGIC SCENARIO
>
> Our son Dani comes home from school and says his friend Y says his other friend X is an S. Later on, we tell Dani he should never say that word again. "What word?", he says. He has not forgotten it, but honestly cannot recall which one of the words he has pronounced we are forbidding him from saying. We feel forced to pronounce "S" in order to make sure he knows what term we are referring to, so we do —we say "We don't call people 'S', that's an ugly thing to say.". We have uttered a slur, even if we have only mentioned it. But we had no other option, and we can be sure that by doing this we have not insulted anyone —if anything, we have prevented Dani from insulting anyone, even if from unintentionally doing so. We know enough about our own son to guarantee that he has understood that we were not insulting anyone. Here, the occurrence of "S" is nonderogatory (Moreno & Pérez-Navarro, 2021, pp. 715-716).

> THE IRONIC SCENARIO
>
> We are a progressive group of friends who would never as much as mention a slur in front of strangers, much less use it to insult a person on grounds of her belonging to a given group. However, we find fun in imitating bigots' mannerisms, and enjoy inner jokes that include ironic uses of "S". We are completely sure that all our friends in the group share our sensibility, and that none of them will take us to aim at insulting anyone. We think it is intuitive to take occurrences of slurs such as these to be nonderogatory, whatever the form of the sentences in which they appear (Moreno & Pérez-Navarro, 2021, p. 716).

The main difference between these two scenarios, MZ&PN claim, is that in THE PEDAGOGIC SCENARIO the slur is mentioned, while in THE IRONIC SCNENARIO, it is used. But, as I mentioned from the outset, and as they also argue, this difference is not relevant when it comes to derogation and thus to producing harm, given that with slurs their derogatory character tends to "scope out" —i.e., hyper-project— both from linguistic and non-linguistic environments.

Now, the next step in their argument is to show that, even with non-derogatory uses of slurs (such as those in the two scenarios presented), there is the potential danger to normalize the use of the slurs in uncontrolled contexts. Thus, the authors show that in relation to THE PEDAGOGIC SCENARIO, by uttering the slur the parents have, inadvertently, given Dani a means to insult the target group —should Dani be so inclined later. Given that the main (or, at least a substantive) harm produced by slurs consists in normalizing a certain way of thinking and speaking about the target group (that is, by normalizing derogation), since the parents have no way of excluding such future uses, even their non-derogatory use of the slur has this normalization potential. The same holds in relation to THE IRONIC SCENARIO: it is entirely possible that, by repeated use, the slur loses its forbidden character and dulls the friends' sensitivity to it, thus making it more likely to be used in uncontrolled contexts. So, to summarize this step in their argument, "no matter how carefully we arrange the current context to make sure that the utterance of a slur does not have the kind of effect we want to avoid, it will facilitate ulterior occurrences of the term. In particular, it will make the slur more likely to appear in uncontrolled contexts in which the utterance of the slur is derogatory" (Moreno & Pérez-Navarro, 2021, p. 716).

Putting all this together, here is what I take their argument to be:

> ***Premise 1***. Normalizing derogation leads to harm.
>
> ***Premise 2***. Uses of slurs in uncontrolled contexts have a high potential to normalize derogation.
>
> ***Partial conclusion 1***: Uses of slurs in uncontrolled context have a high potential to lead to harm.
>
> ***Premise 3***: The potential to normalize derogation is also present with uses of slurs in controlled contexts.
>
> ***Partial conclusion 2***: Uses of slurs in controlled context have a potential to lead to harm.
>
> ***Conclusion***: "[A]ll [uses] of slurs are potentially harmful" (Moreno & Pérez-Navarro 2021, p. 719).

As a final ingredient of their view, and as I already noted, it is important to be clear that MZ&PN are not silentists, as Anderson & Lepore (2013) and others were in the case of, say, slurs used in quotation or embeddings. In other words, while the conclusion of their argument is that all uses of slurs are potentially harmful, this doesn't necessarily support *banning* them across the board. What they hold instead is that each use of a slur comes with a moral cost (as it should be clear by now, *both* in the case of derogatory and non-derogatory uses), and that in some cases "[t]he moral benefits of performing a certain utterance of a slur might outweigh the pervasive moral cost we have described, and so it might be worth it to utter the slur" (Moreno & Pérez-Navarro, 2021, p. 721). Such cases might include the use of the slur in THE PEDAGOGIC SCENARIO, but also uses of slurs in appropriation, in academic contexts,[12] and possibly in the cases pertaining to law and journalism mentioned at the beginning of the paper. As for who bears the responsibility of deciding to use a slur (by carefully weighing the costs and the benefits of using it), MZ&PN are clear that it is the speaker: as they write summarizing their view, "[o]ur point is that uttering a slur always comes at a moral cost, and it is the responsibility of the speaker (...) to assess such cost and decide whether [the slur] is worth [mentioning]." (Moreno & Pérez-Navarro 2021, p. 723).

3. An objection

The Potential Normalization Argument seems to be in good standing, and is consistent with many of the appeals to caution in using slurs non-derogatorily we saw in section 1. Although they rely on two particular scenarios in which slurs are used non-derogatorily, the apport of their argument is more general, and I take MZ&PN to show the limits of a lenient approach to non-derogatory uses of slurs. And even though they don't issue an explicit ban, the point of being cautious, and of thinking of the consequences of such uses, is potently made.

Several objections can be raised to this proposal. To start with, MZ&PN are not entirely clear on the notion of "moral benefits" they are using. Depending on how one understands this notion, a certain consequence of using a slur might count as a benefit or not. Further, since they claim that "performing a certain utterance of a slur might outweigh the pervasive moral cost", this implies that a cost-benefit analysis is performed when a slur is uttered; however, how exactly to perform such an analysis is far from trivial. One could, for example, point out that not all the consequences of a certain act (here, the proffering of a slur) can be known, and so it might be impossible to know whether the benefits outweigh the costs or vice versa. A more general objection is that the view presupposes (or at least comes very close to) a consequentialist outlook: not only has this view been criticised extensively, but there are also many other ethical theories on the market which might apply to the issue of using slurs and possibly yield different results than the consequentialist route the two authors embark on. Finally, a third, more focused objection is that, despite what they assume at the outset, what they show with the two scenarios is that, in fact, there are *no* controlled contexts, leading to the conclusion that the scaffolding they use to support their argument is shaky.

---

[12] The two authors engage in a larger discussion about the effects of slurs in academia, reacting for example to Kukla's (2018) and Herbert's (2017) stances on the matter. As academic contexts are not my focus here, I will not enter that discussion.

However, in what follows I will not push my argument along any of these directions; instead, what I want to focus on is an issue having to do with neutral uses of slurs. The complaint is that, although they consider such uses —which is a step forward in the literature— their reaction to them strikes me as not entirely satisfactory. To be more precise, what I see as problematic in this respect is not the argument *per se*, but one of the accompanying claims they make —namely, that the responsibility of using a slur resides with the speaker.

I start my case by noting that, for some non-derogatory uses to make sense, normalizing derogation is a necessary condition. For example, at least in some cases of appropriated uses, the starting point is to fight back against discrimination, the latter being aided (at least in part) by normalizing derogation of the target group by means of the slur to be appropriated, while the end goal of the process is to normalize the use of the slur without keeping its derogatory meaning. Thus, basically, in appropriation one type of normalization is replaced by another one (the latter devoid of at least *some*, if not all, derogation); this seems to be the chief way of achieving the change in meaning that is required by the goals pursued by the group doing the appropriation. While the range of the group within which such normalization is sought varies (in the case of the n-word, for example, it is the target group, whereas in the case of "queer", the group is everyone), the main goal remains the (arguably gradual) change in meaning. But, quite trivially, for this to happen, normalized derogation by means of the slur has to be in place, with further, repeated, appropriated uses leading to the desired change in meaning. These considerations show two things: that normalization, in itself, is not the (essentially) problematic part in using slurs and that in certain cases normalized derogation is, in fact, a precondition for using slurs in certain ways. But, if this is the case, MZ&PN's argument —which takes the potential normalization of derogation as the main problem with using slurs— loses some of its bite.

A related issue is this. It has been observed by various authors that, in the early stages of appropriation, it is quite difficult to say whether a particular use of a slur is entirely devoid of derogation or still carries it.[13] Additionally, as Brontsema (2004) shows, appropriation is a complex phenomenon than can take many forms, and one dimension of variation is precisely whether the derogatory character of a slur is preserved in appropriation or not (see her comparison between "queer" and the n-word). If these two situations hold, at least for some slurs and in some contexts, then it follows that (quasi-)derogatory uses of slurs are inescapable even in resemantization processes such as appropriation.

Now, MZ&PN's reply is that in appropriation (including the early stages of the process and the types of cases Brontsema focuses on), the benefits of using the slur overweigh the costs, and thus that its use is justified. As we saw above, it is not entirely straightforward how to establish such a claim, given that there is no clear notion of "benefit" the two authors work with. However, even if there were such a notion, the point I want to make is that for other non- derogatory uses of slurs appealing to a "cost-benefit analysis" seems to misrepresent the facts of their very use. For example, a member of a target group using a slur in a neutral way (for example, either using it referentially (in the sense of Anderson, 2018) or in an identificatory manner (as pointed out by Zeman, 2022) seems to happen organically, without the speaker going through a cost-benefit analysis of the sort MZ&PN employ. In the case of the latter, the use of the slur is rather a historical accident, since the target community either didn't have an alternative term or it didn't catch up within the community, and thus it was (or in some cases still is) the only means for some members of the community to identify as part of an ethnic group. This, in turn, casts doubt on the two authors' claim that the responsibility of using a slur resides with the speaker: if using a slur neutrally happens organically without involving a cost- benefit analysis, then it seems exaggerated to ask of speakers from the target communities to pursue one —on pain of being accused of producing potential harm by normalization. Importantly, such uses are crucially different from a derogatory use, which can also happen organically and without involving a cost-benefit analysis (as with people who have been acculturated within a bigoted community), in that it is members of the target group *themselves* that use the slur.

[13] This has led some authors to claim that, at that stage, the slur is ambiguous between a derogatory and a non-derogatory meaning (Jeshion, 2013, 2020; Cepollaro 2020; Jusińska, 2021; Zeman, 2022).

There are two possible replies to this argument.[14] First, the fact that neutral uses happen organically doesn't show that that the speakers from the target group don't go through a cost-benefit analysis. In such cases, the said analysis can be considered to be implicit, or automatic. Second, assuming that speakers using slurs neutrally don't go through a cost-benefit analysis, it doesn't follow that they *shouldn't*: while the former claim is a descriptive/empirical one, the latter is a normative one, and claims of the former sort don't establish claims of the latter sort. In relation to the first reply, one can point out that while it is certainly possible that speakers using slurs neutrally in the spontaneous way mentioned above do, in fact, go through an implicit or automatic cost-benefit analysis, this is something that has to be established by empirical studies in order to be taken seriously. For the likely upshot of speakers going through the analysis in question, even if performed implicitly and automatically, is a significant cognitive burden —or at least one that is greater than that of speakers *not* going through it. As far as I know, no such evidence is available in the case of uses of slurs. Absent such evidence, postulating that the cost-benefit analysis is undertaken by speakers is the more demanding, less economical theory than the one postulating that it is not. So, while I concede the possibility of an implicit and automatic cost-benefit analysis done by speakers in the relevant cases, since it is the proponents of this reply that make this strong claim, it befalls on them to back it up.

In relation to the second reply, while I agree that the descriptive/empirical claim doesn't imply the normative one, it merits stressing that they are connected to each other in a significant way. Suppose one wanted to argue that speakers of neutral uses of slurs should perform a cost-benefit analysis, despite the fact that they do not appear to do so. Certainly, for this claim to be true, these speakers should be capable of performing, or be in a position to perform, such an analysis. There is no doubt that generally these speakers have that ability, given that what it amounts to is merely being reflective and thinking of the consequences of one's words on the well-being of others. However, we have to be careful here. I think that, even if one accepts that the claim that the responsibility of going through a cost-benefit analysis lies with the speaker is a normative one, the case of neutral uses of slurs that happen spontaneously opens up some different, and pressing, *normative* questions. One type of question that appears is about *who* judges that speakers belonging to a certain group should make the cost-benefit analysis and are thus justified or not in using the slur. The social position of those who ask such a thing from speakers (including theorists!) is relevant: if, for example, the speaker is a member of a disadvantaged group, being asked to perform the cost-benefit analysis by a member of an advantaged group raises the threat of *paternalism*: the reality of the life of a member of a disadvantaged group (which might be poverty, lack of education, lack of opportunities, etc.) could simply be erased in the service of casting a general moral judgment. Things might be different when the speaker is part of an advantaged group, for which perhaps the moral stringencies are backed up by access to better living conditions, education, opportunities, etc.

Now, the threat of paternalism doesn't establish that speakers from a disadvantaged group shouldn't be asked to go through the cost-benefit analysis envisaged by MZ&PN, for one could make a distinction between contextual or circumstantial factors that annul a moral obligation and ones that merely make the subject warranted (or blameless) in not fulfilling it. One might thus argue that no circumstantial factors annul a moral obligation or that those mentioned above (poverty, lack of education, lack of opportunities, etc.) are not of the required kind. This touches on delicate ethical issues, and the current paper is not the place to tackle them. The point that I contend myself of making is that whether the threat of paternalism gets in the way of considering asking speakers from disadvantaged groups to go through a cost- benefit analysis before using a slur depends on taking a stance on the ethical issues alluded to. I maintain that this is a substantial burden, one that I'm not sure MZ&PN are ready to take on. In any case, the idea of the social position of those insisting on the cost-benefit analysis seems to me to bring back the issue of the reasonableness of asking speakers from disadvantaged groups to pay close attention to their use of slurs —something that MZ&PN don't fully consider.[15]

---

[14] Both suggested by reviewers, whom I thank.

[15] To better illustrate the type of case I have in mind, let me briefly describe the situation of Roma ethnics in Romania. As many as 650.000 (official number), Roma ethnics have been slaves for several centuries and even after they have been freed, they have continued to live in

A different, less contentious answer to the normative question of who judges that speakers belonging to a certain group should make the cost-benefit analysis and are thus justified or not in using a slur —one that is more sensitive to the social position issue— is this: *other members of that group*. One may argue that other members of a disadvantaged group might be the ones who are better positioned to ask a fellow member to be careful with their use of a slur. Adopting this idea would also explain various intra-group disagreements about whether and how slurs targeting the members of a group should be used by those members *themselves* —as the history of many slurs attests.[16] The difference between this case and the previous one is that the decision of whether and how to use a slur belongs to the target community itself, which should have a say on how language that targets them should be used. Again, as before, and without going into more details, considerations along such lines are not part of MZ&PN's argument itself —but they do seem to affect its plausibility. I think it's fair to say that there is at least a *tension* between the conclusion of the argument, attributing responsibility to the speakers for using a slur, and the neutral uses of slurs focused on.

Now, the two authors address some of these concerns in their papers, having this to say:

> Other nonderogatory uses of slurs that have recently been described are referential (Anderson, 2018) and identificatory (Zeman, 2022). We take these uses to take place in controlled contexts too, as the speaker's group membership is salient enough for her to be confident that the audience will understand that she did not mean to insult, just like happened in Dani's case (Moreno & Pérez-Navarro, 2021, p. 716).

> Referential and identificatory uses of slurs are closer to pedagogic mentions than to ironic uses in this respect. Like with pedagogic mentions, however, there is still the risk that these uses facilitate ulterior occurrences of the term in contexts in which the group membership of the speaker, although salient, does not make the audience understand such occurrences as nonderogatory because the speaker does not belong to the target group (Moreno & Pérez-Navarro, 2021, p. 720).

While these quotes might establish that referential and identificatory uses of slurs have a potential to normalize derogation, they don't address the tension I mentioned above. For even if one considers referential and identificatory uses to happen in controlled contexts, and even if one agrees that they have the potential of being used in uncontrolled ones, there is still no clear answer to the responsibility issue. So, while the conclusion of the Potential Normalization Argument ("all uses of slurs are potentially harmful") remains in place, since the attribution of responsibility goes through a cost-benefit analysis whose reasonability is not unquestioned in cases of neutral uses of slurs, not much follows from it in practical terms. In other words, MZ&PN's analysis doesn't provide an answer to the (normative) question "should *neutral* uses of slurs be banned?", which is essential to answer the larger question of whether all non- derogatory uses of slurs should be banned.

Moreno Zurita (2023) independently considers the two types of neutral uses I marshalled against MZ&PN's view and anticipates some of the issues I raise, as witnessed by the following paragraph:

> Although I have no developed position on this, and only a mere intuition, I do not feel entirely comfortable talking about the harmful potential of appropriated, referential and identificatory uses. I believe that, in these cases, the normalising potential of slurs is irrelevant and that the moral cost assumed, if any for these cases, will always be justified. I just don't think I have

---

poverty. While they are in general better positioned socially and economically in present times, many members of the community still live in suboptimal conditions and are systematically discriminated. The slur used to derogate them, "țigan", has been introduced as an exonym by the ruling classes and is still widely used. Interestingly, however, as discussed in Zeman (2022), a significant number of Roma ethnics use the slur as a way to identify themselves as belonging to the group; such a use is, arguably, non-derogatory. One reason for resorting to the slur for purposes of identification is that some of the Roma ethnics don't speak Romani, their original language, and so have (and had) little choice about the term. Unsurprisingly, I find it unreasonable (and paternalistic) for non-Roma to insist that such members go through a cost-benefit analysis every time they use (non-derogatorily) the term.

[16] See the quick discussion in section 1, subsection 1.2.4.

> the right to say, even if I belonged to one, whether the derogated groups commit any kind of infringement, however minor and justified (Moreno Zurita, 2023, p. 143, footnote 4).

While this is no doubt a morally admirable stance, again, it doesn't get to the heart of the matter. For, on the one hand, the issue is whether there is a cost-benefit analysis present in the case of referential and identificatory uses in the first place, not whether the cost is assumed or the risk of normalization justified. On the other hand, if we take the issue to be a normative one, Moreno Zurita's conclusion seems to be too strong: the problem I pointed out is not that the moral cost is always justified in appropriated, referential and identificatory uses of slurs, but that it depends, in ways unexplored by MZ&PN, on *who* is asking the speakers to police their use. Additionally, saying that in these cases "the normalizing potential is irrelevant" amounts in my view to a serious concession, limiting the main thesis to only a subclass of uses of slurs. This is unsatisfactory because, while derogatory uses are those that best fulfil slurs' main function and are perhaps the most widespread, many uses of slurs are non-derogatory.

4. Summary and conclusions

In this paper, I have engaged with Alba Moreno Zurita and Eduardo Pérez-Navarro's view on the harms of non-derogatory uses of slurs. The two authors urge caution in relation to *any* use of slurs based on their potential —both in what the authors distinguish as controlled and uncontrolled contexts— to normalize derogation. Before getting to their view, I have set the scene by presenting the main types of uses that have been taken to be non-derogatory and the more recent reactions to them in the literature (section 1). The picture that emerged was that non-derogatory uses of slurs can have pernicious effects on the target group, on a par with derogatory ones. Yet, I claimed, due to the variation in the exact attitude towards such uses various authors have recommended, no consensus —ether about a specific type of use or about non-derogatory uses in general— can be extracted, except the rather vague claim that using slurs is permissible under the condition that they lead to some practical or social gain. I turned then to MZ&PN's view, which I take to have the level of generality required.

Against this background, I presented the two authors' main argument —the "Potential Normalization Argument", as I called it (section 2). While I deemed the argument to be in good standing, I have offered an objection to their treatment of neutral uses of slurs (a subspecies of non-derogatory uses) targeting an additional claim they make —namely, that the responsibility for using slurs lies with the speakers (section 3). I have shown that in using slurs neutrally, speakers don't seem to employ the cost-benefit analysis MZ&PN claims should lie at the basis of all uses of slurs, which leads to the claim that taking such speakers to be responsible is not reasonable. I have then considered two replies to this objection, which in turn led to considerations about the role of empirical studies about the use of slurs and about who exactly has the moral standing to demand speakers to make the cost-benefit analysis. However, whether my objection to MZ&PN leads to the rejection of their argument or merely prompts further fine-tuning remains to be seen. While I myself don't offer any consideration for or against banning non-derogatory uses of slurs, I take my contribution to be that of bringing into focus certain (important) aspects not covered by their argument. And while the two authors have moved the issue forward by offering a more general argument than those found in literature and by considering neutral uses of slurs, this (critical) paper is an invitation to think things even further.

Acknowledgments

I thank Alba Moreno Zurita, Eduardo Pérez-Navarro and Sergio Guerra for extensive discussions on the issues tackled in the paper. I also thank the audiences at the *Hate Speech, Fake News and Freedom of Speech* workshop, 18-19.07.2023,

University of Rijeka, the *New Directions in Social Philosophy of Language* workshop, 13-14.11.2024, NOVA University of Lisbon and the *Hybrid Representations* reading group, 10.01.2025, University of Barcelona, where previous versions of this paper were presented. I acknowledge the financial help of a Principal Investigator grant (2023.05952.CEECIND) from Fundação para a Ciência e a Tecnologia - FCT (Portugal).

REFERENCES

Anderson, L. (2015). Racist humor. *Philosophy Compass, 10*(8), 501-509. https://doi.org/10.1111/phc3.12240

Anderson, L. (2018). Calling, addressing, and appropriation. In D. Sosa (Ed.), *Bad words: Philosophical perspectives on slurs* (pp. 6-28). Oxford University Press.

Anderson, L., & Lepore, E. (2013). Slurring words. *Noûs, 47*, 25-48. https://doi.org/10.1111/j.1468-0068.2010.00820.x

Asim, J. (2007). *The N word: Who can say it, who shouldn't, and why*. Houghton Mifflin Harcourt.

Bianchi, C. (2014). Slurs and appropriation: An echoic account. *Journal of Pragmatics, 66*, 35-44. https://doi.org/10.1016/j.pragma.2014.02.009

Black Women's Blueprint. (2016). An open letter from Black women to the SlutWalk. *Gender and Society, 30*(1), 9-14. https://doi.org/10.1177/0891243215611868

Blackburn, S. (1984). *Spreading the word*. Oxford University Press.

Brontsema, R. (2004). A queer revolution: Reconceptualizing the debate over linguistic reclamation. *Colorado Research in Linguistics, 17*(1), 1-17.

Carnaghi, A., & Maass, A. (2008). Derogatory language in intergroup context: Are "gay" and "fag" synonymous? In Y. Kashima, K. Fiedler, & P. Freytag (Eds.), *Stereotype dynamics: Language-based approaches to the formation, maintenance, and transformation of stereotypes* (pp. 117-134). Lawrence Erlbaum Associates.

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

Cepollaro, B., Sulpizio, S., & Bianchi, C. (2019). How bad is it to report a slur? An empirical investigation. *Journal of Pragmatics, 146*, 32-42. https://doi.org/10.1016/j.pragma.2019.03.012

Cepollaro, B., Sulpizio, S., Bianchi, C., & Stojanovic, I. (2024). Slurs in quarantine. *Mind & Language, 39*(3), 381–396. https://doi.org/10.1111/mila.12492

Cepollaro, B., & Zeman, D. (2020). Editors' introduction: The challenge from non-derogatory uses of slurs. *Grazer Philosophische Studien, 97*(1), 1-10.

Croom, A. (2011). Slurs. *Language Sciences, 33*(3), 343-358. https://doi.org/10.1016/j.langsci.2010.11.005

Davis, C., & McCready, E. (2020). The instability of slurs. *Grazer Philosophische Studien, 97*, 63-85. https://doi.org/10.1163/18756735-09701005

Dummett, M. (2007). Reply to Picardi. In B. McGuinness & G. Oliveri (Eds.), *The philosophy of Michael Dummett* (pp. 521-529). Open Court.

Fasoli, F., Paladino, M. P., Carnaghi, A., Jetten, J., Bastian, B., & Bain, P. G. (2016). Not "just words": Exposure to homophobic epithets leads to dehumanizing and physical distancing from gay men. *European Journal of Social Psychology, 46*(2), 237-248. https://doi.org/10.1002/ejsp.2148

Galinsky, A. D., Wang, C. S., Whitson, J. A., Anicich, E. M., Hugenberg, K., & Bodenhausen, G. V. (2013). The reappropriation of stigmatizing labels: The reciprocal relationship between power and self-labeling. *Psychological Science, 20*(10), 2020-2029. https://doi.org/10.1177/0956797613482943

García-Carpintero, M. (2004). The deferred ostension theory of quotation. *Noûs, 38*(4), 674-692. https://doi.org/10.1111/j.0029-4624.2004.00488.x

Gaucher, D., Hunt, B., & Sinclair, L. (2015). Can pejorative terms ever lead to positive social consequences? The case of SlutWalk. *Language Sciences, 52*, 121-130. https://doi.org/10.1016/j.langsci.2015.03.005

Herbert, C. (2015). Precarious projects: The performative structure of reclamation. *Language Sciences, 52*, 131-138. https://doi.org/10.1016/j.langsci.2015.05.002

Herbert, C. (2017). *Exclusionary speech and constructions of community* (Doctoral dissertation). Georgetown University.

Hom, C. (2008). The semantics of racial epithets. *The Journal of Philosophy, 105*(8), 416-440. https://doi.org/10.5840/jphil2008105834

Hornsby, J. (2001). Meaning and uselessness: How to think about derogatory words. *Midwest Studies in Philosophy, 25*(1), 128-141. https://doi.org/10.1111/1475-4975.00042

Jeshion, R. (2013). Expressivism and the offensiveness of slurs. *Philosophical Perspectives, 27*, 231-259. https://doi.org/10.1111/phpe.12027

Jeshion, R. (2020). Pride and prejudiced. *Grazer Philosophische Studien, 97*(1), 106-137. https://doi.org/10.1163/18756735-09701007

Jorgensen Bolinger, R. (2020). Contested slurs: Delimiting the linguistic community. *Grazer Philosophische Studien, 97*(1), 11-30. https://doi.org/10.1163/18756735-09701003

Jusińska, Z. (2021). Slur reclamation – Polysemy, echo, or both? *Organon F, 28*(3), 689-707. https://doi.org/10.31577/orgf.2021.28310

Kennedy, R. (2003). *Nigger: The strange career of a troublesome word*. Penguin Random House.

Kleinman, S., Ezzell, M. B., & Frost, C. A. (2009). Reclaiming critical analysis: The social harms of 'bitch.' *Sociological Analysis, 3*(1), 46-68.

Kukla, R. (2018). Slurs, interpellation, and ideology. *The Southern Journal of Philosophy, 56*(S1), 7-32.

Lasersohn, P. (2007). Expressives, perspective, and presupposition. *Theoretical Linguistics, 33*(2), 223-230. https://doi.org/10.1515/TL.2007.015

Livingstone Smith, D. (2011). *Less than human*. St. Martin's Press.

Livingstone Smith, D. (2020). *On inhumanity: Dehumanization and how to resist it*. Oxford University Press.

Livingstone Smith, D. (2021). *Making monsters: The uncanny power of dehumanization*. Harvard University Press.

Moreno, A., & Pérez-Navarro, E. (2021). Beyond the conversation: The pervasive danger of slurs. *Organon F, 28*(3), 708-725. https://doi.org/10.31577/orgf.2021.28311

Moreno Zurita, A. (2023). *Hiding in plain sight: A non-propositional presuppositional account of slurs* (Doctoral dissertation). University of Granada.

Naylor, G. (1986, March 30). The meanings of a word. *The New York Times*.

Popa-Wyatt, M., & Wyatt, J. L. (2018). Slurs, roles, and power. *Philosophical Studies, 175*, 2879-2906. https://doi.org/10.1007/s11098-017-0986-2

Potts, C. (2007). The expressive dimension. *Theoretical Linguistics, 33*(2), 165-198. https://doi.org/10.1515/TL.2007.011

Rahman, J. (2012). The N word: Its history and use in the African American community. *Journal of English Linguistics, 40*(2), 137-171. https://doi.org/10.1177/0075424211414807

Rappaport, J. (2020). Slurs and toxicity: It's not about meaning. *Grazer Philosophische Studien, 97*(1), 177-202. https://doi.org/10.1163/18756735-09701010

Schlenker, P. (2003). A plea for monsters. *Linguistics and Philosophy, 26*(1), 29-120. https://doi.org/10.1023/A:1022225203544

Smitherman, G. (2006). *Word from the mother: Language and African Americans*. Routledge.

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology, 34*, 379-440.

Stojnić, U., & Lepore, E. (2022). Inescapable articulations: Vessels of lexical effects. *Noûs, 56*, 742-760.

Swanson, E. (forthcoming). Slurs and ideology. In R. Celikates, S. Haslanger, & J. Stanley (Eds.), *Analysing ideology*. Oxford University Press.

Tirrell, L. (2012). Genocidal language games. In I. Maitra & M. K. McGowan (Eds.), *Speech and harm: Controversies over free speech* (pp. 174-221). Oxford University Press.

Williamson, T. (2009). Reference, inference, and the semantics of pejoratives. In J. Almog & P. Leonardi (Eds.), *The philosophy of David Kaplan* (pp. 137-158). Oxford University Press.

Zeman, D. (2022). A rich-lexicon theory of slurs and their uses. *Inquiry, 65*(7), 942-966. https://doi.org/10.1080/0020174X.2021.1903552

Dan Zeman specializes in philosophy of language and formal semantics. He is interested in accounts of various natural language expressions such as indexicals, predicates of taste, aesthetic and moral terms, epistemic modals, slurs, expressives, code words and gender terms. He has published a significant number of papers in journals and collective volumes, and has done editing work as well. He is one of the coordinators of the *Slurring Terms Across Languages* international network.

Address: Institute of Philosophy, Faculty of Letters, University of Porto, Via Panorâmica, s/n, 4150-564, Porto, Portugal. E-mail: zeman@letras.up.pt – ORCID: https://orcid.org/0000-0001-5620-0042