



OPEN ACCESS

EDITED BY

Meg Worley,
Colgate University, United States

REVIEWED BY

Benjamin Lennertz,
Colgate University, United States
Ralph DiFranco,
University of South Dakota, United States

*CORRESPONDENCE

David Miguel Gray
✉ dmgray2@memphis.edu

RECEIVED 07 August 2024

ACCEPTED 20 September 2024

PUBLISHED 20 November 2024

CITATION

Gray DM (2024) How quotation and referential intentions modulate the derogatory force of slur utterances. *Front. Commun.* 9:1477055. doi: 10.3389/fcomm.2024.1477055

COPYRIGHT

© 2024 Gray. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

How quotation and referential intentions modulate the derogatory force of slur utterances

David Miguel Gray^{1,2*}

¹Department of Philosophy, The University of Memphis, Memphis, TN, United States, ²Institute for Intelligent Systems, FedEx Institute of Technology, The University of Memphis, Memphis, TN, United States

The need for new insights to understand the effects of quoting slurs in linguistic communication has been evident over the past several years. Slurs seem to be capable of offending even when embedded in quotations or mentioned. This ability of the derogatory force of slurs to *project* out of embeddings like quotations is an instance of what I will call the ‘projectible force’ of slurs. This force is taken to be a particularly serious problem for content-based semantic theories, which claim that what makes slurring utterances offensive is their derogatory *content*. The inert content criticism claims that quotations should render the content of a slur inert as quotations draw our attention only to the properties of the word itself. However, quoted or mentioned slurs can still offend, so the truth-conditional account of slurs must be wrong. I take the inert content criticism to be a locus of confusion. This confusion can be eliminated if we (1) disambiguate two notions of offense, (2) replace our naïve understanding of quotation with a more precise theory of quotation, and (3) use the deliverances of such a theory of quotation to help explain the different kinds of offense that can result from different features of slurring utterances. While this is not an attempt to defend the truth-conditional account of slurs, I will deliver something that both truth-conditional and prohibitionist accounts of slurs have failed to deliver: an explanation of how slurs under quotation can cause moral offense.

KEYWORDS

slurs, epithets, quotation, mention, use, offense, derogation, hybrid

1 Introduction

One issue that causes a great deal of concern, at least among philosophers, linguists, and anyone who has ever taken a logic class, is the use/mention distinction. This distinction amounts to the following: using a word in an utterance draws attention to how things are in the world. If I say, “The cat is on the mat.” I’m passing along information about a particular cat and what it is on.¹ If I *mention* a word in an utterance, I draw attention to some feature of the word itself. To do this, we often put a word in quotes or italics. For instance, “Canines are

¹ In the sample in the next sentence, ‘the’, ‘is’, and ‘on’ contribute to what the whole sentence picks out in the world, even if they do not individually refer to things (i.e., extensions).

mammals, but ‘canines’ is a plural noun.”² I will use ‘quotation’ very loosely to cover written and spoken cases of mention.

What keeps this academic concern from remaining academic is media coverage of public slur utterances and the obfuscation of the use/mention distinction. For instance, many headlines claim that someone *used* a slur when they, in fact, *mentioned* the slur.³ While there are plenty of instances of people using slurs that make the news,⁴ the cases that tend to capture public attention the most are mentions of a slur. In such cases, it appears that the primary intention of the speaker was *not* to derogate the group that a slur targets, but to draw attention to the word.

Both the desire not to offend and undoubtedly the desire not to be publicly condemned have resulted in some flat-footed application of a ban on the utterance of slurring terms. For instance, a Wisconsin school district created a zero-tolerance policy on racial slur utterances that resulted in the firing of an African American school security guard for telling a student not to call him ‘n*gger’ (a mention of a racial slur in which he admonished the student for using it against black people, of which he was a group member).⁵ Fortunately, we are seeing nuance emerge in the public sphere, mostly from philosophically or linguistically trained public intellectuals.⁶

No doubt uttering or writing slurs, whether used or mentioned, can result in hearers taking offense. Whether slurs can derogate (a feature of a slur utterance) both when used and mentioned is an open question. And, as there is no reason to offend when it can be avoided, a common practice of neither using nor mentioning slurs is emerging in the media and academic literature. In the latter case, this is evinced by the recent practice of papers in the slur literature of replacing a few letters with asterisks (as I will do) or referring to slurs with phrases—such as ‘c-word,’ ‘b-word,’ and ‘n-word.’

Determining what is wrong with mentioning or quoting slurs, if anything, is also a concern for developing an adequate linguistic and

philosophical account of slurs. The apparent ability of the derogatory force of slurs to *project* out of embeddings like quotations is an instance of what is called an ‘embedding failure’ (e.g., Bolinger, 2017, p. 439) and what I will call the ‘projectible force’ of slurs.⁷ This phenomenon is something theories of slurs need to explain or explain away.

One of the most prominent theories of slurs claims that the derogative powers of slurs are a feature of their semantic truth-conditional content (Hom, 2008). Such theories hold that the source of offense for slurs lies in their semantic content: a slur means that a person or group is deplorable or worthy of contempt in virtue of being a member of the targeted group.

There is a standard take on truth-conditional accounts of slurs and their purported inability to explain the projectible force of mentioned or quoted slurs. I’ll call this ‘the inert content criticism.’ Here is the condensed version, which I will expand upon in Section 3. Mention or quotation of a slur draws our attention to the word itself. As such, the ‘content’ or ‘meaning’ of the slur is rendered inert. And if the content is rendered inert, the derogatory features of the content are also rendered inert. Yet, mentioned or quoted slurs, as exhibited by the media examples, still seem to have the capacity to offend by virtue of being uttered. As such, a truth-conditional account of slurs cannot be correct.

I do not think this argument works for several reasons. However, my goal here is not to launch a defense for a truth-conditional account of slurs. Rather, I think the argument suffers from several unclaritys; unclaritys which are pervasive in the literature on slurs. Focusing on attempts to explain the offensiveness of slurs under quotation from both truth-conditional and non-truth-conditional perspectives will help reveal some of the problems with this argument. From this investigation, I plan to argue for the following theses:

- 1 Slur utterances can cause at least two kinds of offense: ‘moral’ and ‘affective.’ Not disambiguating these two kinds of offense hides a problem with some accounts of offense.
- 2 It is false that quotation renders content, let alone derogatory content, inert. This idea comes from using an underdeveloped notion of quotation that can be remedied by appealing to a theory of quotation that covers all types of quotation. I’ll make use of such a theory provided by Saka (1998, 2003).
- 3 Appealing to the above theory of quotation allows us to precisify what it is about truth-conditional content that allow a slur utterance to be the source of moral and affective offense, even in cases of quotation.

I will proceed as follows. Section 2 will highlight some distinctive features of slurs, present some of the linguistic mechanisms that theorists have posited are sources of offense, and some of the explanatory challenges for theories of slurs that are relevant to the current project. In section 3, I will present the inert content criticism in detail. This presentation will help clarify what features of the

2 As a convention, I will use single quotation marks when mentioning a word or phrase. If I have to mention a word or phrase within another quotation, I will use double quotation marks for the larger quotation.

3 For instance: <https://www.nbcnews.com/news/us-news/duquesne-university-professor-who-used-racial-slur-class-put-leave-n1239957>, <https://fox17.com/news/local/ut-students-call-for-professors-resignation-after-viral-tweet-of-racial-slur-university-of-tennessee-knoxville>, <https://www.insidehighered.com/quicktakes/2020/10/08/duquesne-dismisses-professor-who-used-racial-slur-class>, <https://bwog.com/2021/04/professor-dinah-pokempner-uses-racist-slur-over-zoom/>.

4 Three recent cases include <https://www.cnn.com/2021/12/16/us/louisiana-lafayette-judge-racist-language-video-michelle-odinet/index.html> and <https://www.courant.com/coronavirus/hc-news-coronavirus-southington-racial-bias-arrest-20211029-kv6ymg2nfpzppsf4yrcnrluxi-story.html>, and <https://www.newsweek.com/nj-towing-company-loses-government-contract-after-worker-uses-n-word-camera-1622309>.

5 He was eventually rehired after a petition. <https://time.com/5706848/black-security-guard-wisconsin-n-word/>

6 For instance, linguist John McWhorter (e.g., <https://www.nytimes.com/2022/02/11/opinion/use-mention.html>), linguist Caitlin Green (e.g., <https://c-moriah-green.medium.com/beyond-mention-vs-use-the-linguistics-of-slurs-3e0bfff1c5d>), and philosopher Kwame Anthony Appiah (e.g., <https://www.nytimes.com/2022/04/05/magazine/racial-slur-classroom-ethics.html>).

7 What the derogatory force of a slur seems to be projecting out of are particular linguistic embeddings such as quotation, negation, the antecedents of conditions, etc. As such it is also said that slurs exhibit an embedding failure evinced by slurs still being offensive in such linguistic contexts.

criticism I will be objecting to and addressing. One problem with the inert content criticism is that it equivocates between two kinds of offense. In Section 4, I will distinguish between what I call ‘affective’ and ‘moral’ offense and present three reasons for keeping them distinct (Section 4.2.1).

To address the inert content criticism’s claim that quotation renders content inert, we will need to examine what quotation is as well as what an adequate account requires. In short, an adequate account of quotation needs to cover all kinds of quotation, including scare quotes, and provide us with an intuitive understanding of the use/mention distinction. I will draw heavily from Paul Saka’s work to this effect (1998, 2003). This theory of quotation and its application will be the topic of Section 5. The result of this investigation will show that content is not rendered inert by quotation. Rather, the aspect of quotation that prevents slurs from derogating is that the speaker does not draw the audience’s attention to the intended referent of the term. However, I’ll argue that in the case of direct quotation, *moral* offense can still result in virtue of the quoted speaker making referential use of a slur.

With this disambiguated account of offense and clarified account of quotation, I will present three approaches to theories of slurs in Section 6. The first account posits the sociolinguistic mechanism of prohibition violation in accounting for the offensiveness of slurs, whether used or mentioned (Anderson and Lepore, 2013a, 2013b). I will argue that this theory only provides an account of affective offense. The second and third accounts (Hom, 2008; Rinner and Hieke, 2022, respectively) are defenses of truth-conditional accounts. Hom seems to deny that slurs under quotation cause offense, while Rinner and Hieke claim that slurs under quotation do offend. I’ll conclude that both views end up demonstrating the same thing: that slurs under quotation can cause affective, but not moral offense. In contrast to these three theories, my account shows the conditions under which mentioned and quoted slurs can cause moral offense and why all slur utterances can cause affective offense.

Four caveats should be mentioned at this time. First, for reasons of length, I will not be examining the majority of theories of slurs, in particular pragmatic theories. I do not claim that my concerns about content-based theories of slurs can generalize to pragmatic theories. This claim is a topic for another investigation. What I will do is contrast a truth-conditional account with an account that seems best equipped, if not custom-made, to account for the derogatory force of slurs in embedded contexts. Second, some theorists focus on the derogatory features of slurring *terms*, while others focus on the derogatory features of slurring *utterances*. This difference can result in confusion, especially when considering the content-based theories of slurs. While the content of a slur can be derogatory, content itself does not derogate. Rather, utterances of slurs can derogate, with the clearest cases being when slurs target groups or individuals. I will take this to be understood when I talk about some theories claiming that the derogatory force of a slur is part of its semantic content. Third, I make use of a particular theory of quotation primarily because it provides an account for all purported kinds of quotation, which I take to be a good-making feature of a theory of quotation. I also find the theory highly plausible and intuitive. However, for reasons of length, I will not provide a defense of this theory against competing theories. Fourth, this is not a defense of the claim that all the derogatory force of a slur necessarily comes from the truth-conditional content of slurs. I think there can be multiple linguistic and non-linguistic sources of

derogatory force that can contribute to moral and affective offense, although I will not be arguing for that position here.⁸ What I will argue is that there are serious confusions around the ideas of projectible force, taking offense, and what is communicated in quotation. The aim of this article is to remedy those confusions.

2 Slurs and theories of slurs

2.1 Locating slurs among other negatively valenced terms

Slurs can be contrasted with other terms that have a negative valence. Some subtypes of negative valenced terms are *swear words* (e.g., ‘shit,’ ‘fuck’), insults (e.g., ‘asshole,’ ‘jerk’), negative-factive descriptors (e.g., ‘thief,’ ‘criminal,’ ‘killer’), and slurs (e.g., ‘h*eb,’ ‘r*dneck,’ ‘Sp*c’). While insults normally target individuals, negative-factive descriptors and slurs are often used to target both individuals and groups.

Slurs normally target groups or individuals whose group properties are of social significance. Examples of such group properties are nationality, immigrant status, geographical origin, race, ethnicity, gender, sexual orientation, religion, ability status, or common aptitudes (Hom, 2010, p. 165; Anderson and Lepore, 2013a, p. 25; Jeshion, 2013, p. 232; Bianchi, 2014, p. 35; DiFranco, 2017, pp. 372–73; Nunberg, 2018, p. 239; Gray, 2024). A ‘standard use’ or ‘targeted use’ of a slur involves a competent language user uttering a slur with the intention of picking out the group or its members typically associated with the slur. Standard uses are linked with malevolent intentions, such as the intent to insult. I do not have a comprehensive list of non-standard uses. Still, some examples include utterances of a slur when it is embedded in a quote (pure or direct quotation), many instances of indirect quotation, and possibly corrective and educational uses (Hornsby, 2001, pp. 128–129; Hom, 2008, p. 429). These are contexts where the intent of the speaker may not be malevolent.

One main difference between slurs and other negatively valenced terms is that standard uses of slurs seem always to be unwarranted (Gray, 2024). The reason is that slurs derogate members of a group by virtue of their group membership. Warranted derogation would require that some constitutive feature(s) that determine group membership are morally apt.⁹ For instance, if thievery were a feature that constituted group membership, then deriding a group for having that feature would be permissible. Admittedly, figuring out what features constitute racial groups is a complicated issue.¹⁰ However, on the variety of accounts of what races are, none of them hold that morally apt properties are constitutive features of race. For

8 I argue for the existence of non-linguistic sources of offense and suggest a few reasons to take a pluralist approach to sources of derogation in Gray (2024).

9 By ‘morally apt’ I mean a property that is relevant for moral assessment such as being a murderer.

10 Works on the metaphysics of race are legion. Most current views on race take them to be social categories. For some recent socially-based theories of race, see Mallon (2016,2022), Ásta (2018,2022), Glasgow et al. (2019), Haslanger (2020), Gray (2022a), and Hardimon (2022). For an overview, see Gray (2022b).

instance, no racial group has, among its defining properties, features such as a propensity towards thievery or malevolent actions. And since the properties that make up these groups are morally inapt, contempt or derogation based on these properties would be unwarranted. (The same would hold for other kinds of groups, such as those mentioned above.) The unwarrantedness of slur utterances is one way to capture why hearers are warranted in taking offense when they hear slur utterances in standard or targeted use cases (Gray, 2024).¹¹

Having contrasted slurs with other negatively valenced terms, we can make the following claims.

- 1 Slurs are negatively valenced.
- 2 A standard or targeted use of a slur is directed towards groups or individuals insofar as they are members of the group.
- 3 The basis for actual group membership is a set of properties that are morally irrelevant and thus cannot warrant derision.

2.2 Theories of slurs and sources of offense

I mentioned earlier that the phrase ‘derogatory force’ is a term of art that refers to the derisive power of slur utterances originating from a semantic or pragmatic mechanism. However, some theorists of slurs have posited linguistic mechanisms that can cause offense that are not semantic or pragmatic in nature.¹² I will thus use the phrase ‘derogatory force’ a bit more loosely to cover other linguistic mechanisms. Additionally, I will use ‘derogatory force’ interchangeably with the less loaded term ‘sources of offense’ to encompass any speaker-side linguistic mechanism that can be the source of offense for a hearer.

Among the posited linguistic mechanisms that can act as a source of offense in slur utterances are phonetic features (Mandelbaum et al., 2024), truth-conditional content (Hom, 2008; Sennet and Copp, 2015; Bach, 2018; Neufeld, 2019), the expressive attitudinal component (Potts, 2007; Saka, 2007; Boisvert, 2008; Richard, 2008; McCready, 2010; Croom, 2011; Gutzmann and Gärtner, 2013; Jeshion, 2013), conventional implicature (Williamson, 2009; McCready, 2010; Whiting, 2013; Lycan, 2015), presuppositional content (Schlenker, 2007; Predelli, 2010; Cepollaro, 2015; Cepollaro and Stojanovic, 2016), inferential role (Dummett, 1973; Tirrell, 1999), meta-content (Blakemore, 2015), conversational implicature (Bolinger, 2017;

Nunberg, 2018), and the sociolinguistic mechanism of prohibition/taboo (Anderson and Lepore, 2013a).

2.3 Explanatory desiderata

Providing an account of the source of offense is normally done in the service of answering other pressing questions posed by the use of slurs, which we may call ‘explanatory desiderata.’ Accounting for the offensiveness of slurs is not the only thing a theory of slurs ought to explain. However, it is a central concern, and many of the desiderata focus on the offensiveness of slurs. Most theories of slurs try to provide explanations of the desiderata by virtue of the linguistic mechanism the theory takes to be the source of offense. My goal is not to defend such a theory, nor is it to respond to all of these desiderata. Rather, I want to draw attention to a few of the several desiderata as they reveal important features of slurs and problems for slur quotations and mentions. For these purposes, it will be useful to present four explanatory desiderata:

Derogatory Autonomy: Slur utterances act as a source of offense independently of a speaker’s intentions (Tirrell, 1999; Hom, 2008, 2010; Anderson and Lepore, 2013a; Jeshion, 2013; Bolinger, 2017; Liu, 2021).

Offense-Taking Autonomy: Slur utterances cause offense to be taken independently of an utterer’s intentions (Anderson and Lepore, 2013a; Jeshion, 2013; Bolinger, 2017; Popa-Wyatt and Wyatt, 2018; Liu, 2021; Rinner and Hieke, 2022)¹³

Projectible Force: The derogatory force of slurs project out of a variety of embeddings—such as negations, antecedents of conditionals, modals, and pure, direct, and mixed quotation (Hornsby, 2001; Potts, 2004, 2007; Hom, 2008, 2010; Richard, 2008; Anderson and Lepore, 2013a; Camp, 2013; Jeshion, 2013; Bolinger, 2017; Rinner and Hieke, 2022).¹⁴

Synchronic Derogatory Intra-variation: Why can the same slur, in the same period, vary in derogatory force in different linguistic

11 It might be claimed that uses of slurs could be warranted (e.g., see Difranco and Morgan, 2023). For instance, we might imagine coming up for a slur for ‘whiteness’ that captures the unjust status of whites on top of a social hierarchy. Difranco and Morgan suggest “If it is possible for a derogatory attitude to be morally just, then morally permissible (and even praiseworthy) slurring acts may be possible as well” (Difranco and Morgan, 2023, p. 486). While I cannot fully address this view here, we can note a difference between these and other slurring terms. Slurs are easily applied to individuals and groups in virtue of a part-whole relationship. However, if it features worthy of derogation are systemic features, then such a term is less easily applied to individuals without modifying the meaning of the term.

12 I am using ‘linguistic mechanisms’ broadly to cover sociolinguistic mechanisms in addition to semantic and pragmatic mechanisms.

13 The desideratum of Derogatory Autonomy involves providing an explanation what mechanisms allow slur utterances to derogate or be a source of offense independently of the speaker’s intention. This is a ‘speaker-side’ phenomenon. Offense-Taking Autonomy involves providing an explanation of how slurs can result in an audience taking offense upon hearing a slurring utterance, regardless of the speaker’s intentions. This is a ‘hearer-side’ phenomenon, albeit caused by the speaker’s utterance (see Section 4.1 for more details). Much of the slur literature does not distinguish derogatory autonomy from offensive autonomy (but see Liu, 2021). One piece of evidence for this is that almost any author that provides such a list does not list them as distinct desiderata (e.g., of those who provide lists of explanatory desiderata, see Hom, 2008; Nunberg, 2018; Potts, 2007).

14 Others talk of non-displaceability (Potts, 2004; Hom, 2010; Anderson and Lepore, 2013a) or embedding failures (Jeshion, 2013). For an illuminating discussion of embedding or projection Asher (1993), Potts (2004), and Hunter and Asher (2016).

contexts (Bolinger, 2017; Nunberg, 2018; Popa-Wyatt and Wyatt, 2018; Cepollaro et al., 2019).

Now that I have presented what I take slurs to be, some important explanatory desiderata, and a list of linguistic mechanisms that are posited as the source of a slur utterance's offensiveness, I will revisit the inert content criticism in more detail.

3 The inert content criticism

Searle captures an important point about mentions and quotations when he says of quotation, "The word itself is *presented* and then talked about, and that it is to be taken as presented and talked about rather than used conventionally to refer is indicated by the quotes" (Searle, 2011, pp. 75–76). To indicate mentioning in writing, we normally use quotation marks, italics, or at least an introductory phrase such as 'the word.' We can make Searle's point more concrete by seeing how certain predicates render quoted or unquoted subjects true or false. Take the following true sentences:

[1] Cats are mammals.

[2] 'Cats' has four letters.

In [1], the subject 'cats' is used to talk about cats, and the predicate 'are mammals' assigns the property of being a mammal to cats. In [2], the subject "'cat'" is *presented*, and the predicate 'has four letters' assigns the property of having four letters to the word 'cat.'

I have not found any explicit argument for why the derogatory force of truth-conditional content is rendered inert. Nevertheless, it is taken to be a standard problem for truth-conditional accounts of slurs. For instance, Anderson and Lepore claim that "quotation famously renders content inert" but that mentions of slurs "can easily cause alarm and offense" (2013a). Bolinger makes a similar point noting the embedding failures of slur utterances in cases of mention and indirect reports (2017, p. 439). Rinner and Hieke provide a formulation of the inert content criticism in their attempt to defend content theories of slurs (2022, p. 1485).

We can combine the important points made in these claims, along with some presuppositions, to create the following argument.

The Inert Content Criticism:

- 1 An adequate theory of slurs must explain how slurs can cause offense when they occur in mentions or quotations, as this is part of explaining the projectible force explanatory desideratum.
- 2 Truth-conditional content theories of slurs claim that the source of offense in slurring utterances is a feature of their truth-conditional derogatory content.
- 3 Quotation/mention renders content inert.
- 4 A truth-conditional theory of slurs cannot explain how slurs cause offense in mentions or quotation [from 2 and 3].
- 5 A truth-conditional theory of slurs cannot explain the projectible force explanatory desideratum [from 1 and 4].
- 6 Therefore, truth-conditional content theories of slurs are not adequate theories of slurs. [from 1 and 5].

The inert content criticism still leaves us with several questions. Namely, what do we mean by offense, what does it mean to say that

the offensiveness of slurs projects out of quotation, and why think that quotation renders content inert? I will address the first of these questions in the next section.

4 Sources of offense and offense taken: some distinctions

4.1 Speaker-side source of offense vs. hearer-side offense taken

First, as already noted, it is important to keep in mind the distinction between the speaker-side source of offense (commonly referred to as 'derogatory force' when such force is taken to be a semantic or pragmatic feature of an utterance) and the hearer-side phenomenon of taking offense. The term 'offensive' most naturally applies to a feature of a slurring utterance. To be offensive is to have the capacity to cause offense. This means that taking offense is often our guide to what is offensive. However, people may not take offense when something offensive has occurred, and people occasionally take offense when there is not a good reason for it. This suggests a need for caution when offense is taken, as it is still possible that nothing offensive has occurred. But suggesting that offensiveness and taking offense does not always track one another also demands an explanation. What is needed is more clarification on the nature of offense taken in order to determine what are reasonable guides to an utterance being a source of offense.

4.2 Affective and moral offense taken

If we look at work in both psychology and philosophy, it becomes clear that there are at least two kinds, or at least aspects, of offense. I take it they can occur independently of one another, although their co-occurrence is common, leading to some of the confusion discussed above.

We can call one type of offense 'affective offense.' Often, being offended has affective dimensions. The affect of offense is often associated with emotional processes and its effect on cognitive processing (Tamás et al., 2010; Madan et al., 2017; White et al., 2017; Hansen et al., 2019). By 'affect,' I have in mind what researchers in the psychology of emotion have in mind:

In the science of emotion, "affect" is a general term that has come to mean anything emotional. A cautious term, it allows reference to something's effect or someone's internal state without specifying exactly what kind of an effect or state it is. It allows researchers to talk about emotion in a theory-neutral way (Barrett and Bliss-Moreau, 2009).

For instance, the presentation of taboo words (which include but are not limited to racial epithets) is known for causing high states of arousal, which is partially constituted by a heightened response from the autonomic nervous system (Janschewitz, 2008). This autonomic response is associated with wakefulness, increased blood pressure and heart rate, sensory alertness, attentiveness, and a readiness to respond (Madan et al., 2017; White et al., 2017). Thus, affect is at least partially explained by physiological features.

In contrast to affective offense, we can call the other kind of offense ‘moral offense.’ This is the offense that one can reasonably take towards wrongful conduct. If a speaker has done something wrong by uttering a slur and the hearer is made aware of this wrong by virtue of the utterance (whether experienced directly or indirectly), then the hearer can take moral offense. (More precisely, we could call this offense ‘morally justified offense’ as the offense requires the offended to have more than *prima facie* justification to think the utterance is the source of moral offense).¹⁵

I have not encountered the distinction I am making in the literature on slurs, nor does anyone seem to be making use of this distinction.¹⁶ I do not take these categories to be new, but they have not been distinguished in the literature on slurs. I will further clarify the distinction between these two kinds of offense in three ways.

I will say that in the Philosophy of Law, [Feinberg \(1987\)](#) makes distinctions between aspects of offense, some of which map onto my distinction here. In his account of ‘profound’ offense, he defines the specific normative aspect of taking offense as being caused by the wrongful conduct of others. Some of the features he assigns to profound offense are features of what I call ‘moral offense.’ Profound offense is ‘serious’ and can occur without the offensive conduct being experienced directly (e.g., if one knew that there was a white supremacy rally in your hometown, but you were not there to see or hear it). Additionally, profound offense affects us as moral agents as the conduct “offends because it is believed to be wrong, not the other way round. It is not believed to be wrong simply and entirely because it causes offense” ([Feinberg, 1987](#), pp. 58–59). This claim is important as it highlights that being offended does not make something wrong. While experiencing offense can still be a guide to moral offense, merely being offended does not mean a moral wrong has been committed. Finally, profound offense is at least partly impersonal. Being offended at the desecration of religious icons, the flagrant display of white supremacist symbols, or the loss of civilian life in warfare are offensive not just because they affect us as individuals, but as groups—religious, racial, or human.

4.2.1 Offense: different concerns, causes, and importance

To dissuade possible concerns that I have created a distinction without a difference, there are three reasons we can and should, distinguish affective and moral offense. The first reason is that we care about affective and moral offenses in very different ways. Regarding affective offense, most of us do not want to cause anyone physiological or emotional discomfort, and we will often take several steps to make sure this does not happen. However, in some cases, creating a negative affect is taken to be acceptable for the achievement of other goals. For instance, end-of-year exams, teaching children to swim, and

life-saving surgeries cause emotional discomfort that we find acceptable. Activists have even subjected themselves to the emotional discomfort of slurs. In preparation for the Freedom Rides of 1961 (bus trips taken by civil rights activists to test recent desegregation laws in southern bus terminals), the student-founded CORE (Congress for Racial Equality) ran non-violent training sessions to ensure participants could remain non-violent. This included subjecting potential riders to uses of slurs, other demeaning language, and mild physical abuse.¹⁷ However, in the case of moral offense we are more concerned not just with discomfort or harm (which can be morally permissible) but also with the fact that a moral wrong has occurred.

The second reason for distinguishing affective and moral offenses is that they relate to slurring utterances in different ways. For instance, when first learning that a particular word is a slur, one might understand that there is a reason to take moral offense, even if no affective offense is present. Once the slurs are understood, standard use cases of slurs can cause strong affective responses. Over a brief amount of time, the affective response could be causally associated with a slur utterance, even in non-standard use cases. In some cases, hearing a term that sounds superficially like a slur could cause affective offense by virtue of associative learning—even if one knows that taking moral offense is inappropriate. By contrast, moral offense cannot be taken in virtue of a slur utterance unless one knows that certain conditions have been met. For instance, if the slur utterance counts as a standard or targeted usage of a slur, it commits a moral wrong and moral offense can be taken.

I should note that I am here concerned with cases when the moral wrong occurs in virtue of some linguistic mechanism. However, moral offense can occur via, but not in virtue of, a linguistic mechanism, even in non-standard uses. For instance, we can imagine an educator talking about slurs to a population who have been significantly harmed by a particular slur such that the harm of hearing it outweighs the educational benefits of the slur utterance. Racist wrongs can occur not only through malevolence but also through negligence. If the speaker has not familiarized themselves with the population and profusely mentions the slur, the speaker can commit a moral wrong. But the wrong is not one of derogation. Rather, it’s one of not showing enough concern for those that are being communicated with.

The third reason to distinguish between affective and moral offense is that moral offense is the case philosophers and linguists are primarily concerned with.¹⁸ Standard use cases of slur utterances are

15 *Prima facie* justification that could be defeated would be something like taking offense upon hearing a word by a foreign language speaker that sounds like a slurring term in one’s own language. This could cause an immediate affective offense, and one may initially believe they also have reason to be morally offended.

16 This is not to say this is the only way to make helpful distinctions between different kinds of offense. For example, [Bolinger \(2017\)](#) distinguishes between actual, rational, and warranted types of offense.

17 I take the use of slurs in these cases not to be morally offensive for two reasons. First, the exercise CORE conducted was a kind of play acting and should be taken to be more like a fictional use of slurs in an, otherwise, well-intentioned book or play. Second, we may be able to invoke the Doctrine of Double Effect in this case, which holds that it is morally permissible to perform an action that can harm an individual when that action is a side-effect (not the primary intention of the person) of an action where the benefits significantly outweighed the harms. In this case using a slur could be permissible (and thus not morally wrong). This is an extremely flat-footed account of what ethicists call the ‘doctrine of double effect’. For a more nuanced account see [Connell \(1967\)](#), [Marquis \(1991\)](#), and [Wedgwood \(2011\)](#). For a criticism of the doctrine, see [Scanlon \(2010\)](#).

18 More precisely, we are concerned with a linguistic analysis of what features of slurs are responsible for wronging an individual or group. In addition to

particularly adept at communicating unwarranted contempt. However, to derogate an individual or group based on group properties that are not apt for negative moral evaluation is to commit a moral wrong.¹⁹ It is warranted moral offense, not merely affective offense, that has created an interest in the linguistic mechanisms that allow the communication of such contempt. Regarding this function of slurs, Jeshion claims that “Slurs function to derogate or dehumanize, by which I mean, that they function to signal that their targets are unworthy of equal standing or full respect as persons, that they are inferior as persons” (2013, p. 232). What we see in Jeshion’s account and those of others is that when we attempt to account for the source of offense in slur utterances—whether in terms of semantic, pragmatic, or sociolinguistic mechanisms—the offensiveness we care about cannot just be affective.

I take this all to suggest that at least two distinct concepts fall under the umbrella term ‘offense’ and that distinguishing between them is of importance in discussing slurs. More importantly, if we take it that slur utterances derogate or show contempt, then any explanation of the offense taken should be able to explain moral offense, if not both moral and affective offense. Thus, in explaining cases of moral offense taken in the presence of slur utterances, whether used or mentioned, we need to be able to explain why slur utterances wrong.

5 A theory of quotation

5.1 Kinds of quotation

There are some central assumptions made concerning quotation in work on slurs. I take two assumptions to be that (1) if a slur utterance has derogatory content, then quotation or mention ought to render that content inert, and (2) The first point is true of cases of pure and direct quotation. The first point is explicit in the literature. The second point I draw from the types of quotation that are normally given in the literature either two challenge or defend the idea that quotation renders derogatory content inert (e.g., see [Hom, 2008](#); [Anderson and Lepore, 2013a](#); [Bolinger, 2017](#); [Rinner and Hieke, 2022](#)) This second point does not deny that other types of quotation might act in similar ways.

Given the importance placed on the issues of projectible force regarding slurs under quotation, it may seem surprising that the substantial literature on quotation is almost never used.²⁰ One reason may be that there is little agreement on what quotation is, what counts as cases of quotation, and what the explanatory desiderata are regarding quotation. I cannot hope to settle such issues here. I will point out a good-making feature of a theory of quotation.

wronging individuals, slur usage can have a cumulative effect, perpetuating social injustices. This is not to say that affective offense is of no interest. Psychologists may choose to study affective response because they are interested in the physiological features and the effects affective response has on us over time. For similar reasons a few articles on slurs do focus explicitly on this feature (e.g., [Mandelbaum et al., 2024](#)).

¹⁹ In this way, slurs can be contrasted with insults like ‘asshole’ which are usually based on morally evaluable group properties and can be warranted.

²⁰ For an excellent introduction to philosophical and linguistic issues concerning quotation, see [Cappelen et al. \(2023\)](#).

First, it is important to appreciate the different uses of quotation.

Direct Quotation is a verbatim report of someone’s utterance.

Indirect Quotation is a non-verbatim report of what someone said, although it might reuse important terms.

Mixed/Hybrid Quotation is a partially verbatim report of what someone said. The verbatim portion is both used and mentioned. In cases of mixed quotation, the speaker makes use of the utterance that someone said in explaining what they said.²¹

Pure Quotation does not present what someone has said but merely mentions words, phrases, or sentences.

Scare Quotes are treated by some as mixed quotation but are often not attributed to a speaker. Scare quotes can be used to signal different, often worrisome, features of a term such as that it hasn’t been well defined or that it is not the best term to use for a given purpose. A phrase such as ‘so-called’ can be used in some cases of scare quotes in speech to signal that scare quotes are intended.

Examples of these kinds of quotation are:

[3] *Direct Quotation*. Maria speaking: Jane said, “When it comes down to it, I do not trust Sp*cs.”

[4] *Indirect Quotation*. Maria speaking: Jane said that she does not think Sp*cs are trustworthy.

[5] *Mixed/Hybrid Quotation*. Maria Speaking: Jane said, “When it comes down to it” he does not trust Hispanics.

[6] *Pure Quotation*. “Cat” is a count noun.

[7] *Scare Quotes*. The local online ‘experts’ are saying we should not wear masks.

A good-making feature of a theory of quotation is that it is a theory of all kinds of quotation. If a theory can provide a unified account of these phenomena, all the better for the theory of slurs. While not all theories attempt to account for mixed quotation and scare quotes, perhaps claiming that they aren’t really kinds of quotation, a notable feature of Paul Saka’s account of quotation is that he explains all types of quotation (1998, 2003).

²¹ While mixed quotation may strike some as an oddity, as attention was drawn to it quite recently by [Davidson \(1979\)](#). However, its use is ubiquitous in news headlines. For instance, “A lack of trust”: How deepfakes and AI could rattle the US elections” ([Beaumont, 2024](#)); “Curtail free speech? ‘Oh yes we can” ([Halimi and Rimbart, 2024](#)). Among the debates regarding mixed quotation is whether there is a distinction between scare quotes and other cases of mixed quotation, as they both involve elements of use and mention. For an overview, see [De Brabanter \(2010\)](#). An important difference is that scare quotes need not be attributed to a speaker.

5.2 Saka's disambiguation theory of quotation

Saka (1998) claims that speech acts (meaningful verbal utterances, normally used for communicative purposes) are ambiguous in that they have multiple 'ostensions.' To 'ostend' is to show or point out something intentionally in speech or writing (Assimakopoulos, 2022, p. 46). For Saka, if one is speaking or writing, one both makes manifest several features associated with what one says and can intend to do so in a way that is recognizable to whomever one is communicating with (2003, p. 187). 'Direct' ostension involves pointing out something immediately present to one's audience (such as the sight or sound of a word), whereas 'deferred' ostension involves drawing an audience's attention to something that is not directly present by virtue of something that is (e.g., pointing to a car with a parking ticket and saying 'he's going to be upset'). Thus, for Saka, accounting for use and mention requires that use or mention directly and deferringly ostend several features of a word or phrase. Using Saka's example of the word 'cat,' we can detail the process by which we come to understand a speaker's use of the word. An utterance of 'cat' automatically ostends to its audience the lexeme (i.e., an abstract unit of meaning) for 'cat.'²² Among the lexical features that are ostended are the phonetic form /kæt/, the syntactic category noun, the intension (or concept) CAT,²³ and the customary referent (or extension) of CAT ({x: x a cat}) (Saka, 1998, p. 126).²⁴ In later work, Saka puts this claim of intension (or concept) in terms of content when he states, "In uttering any expression *x*, S defeasibly intends for the audience to execute *x*'s lexico-syntactically specified conceptual content" (2003, p. 190).

There is debate concerning how to best construe semantic content. However, that content determines reference is an appealing feature for a semantic theory in that a meaningful utterance is assigned a value, or content, that determines the referent for that utterance.²⁵

The use or mention of an expression is multiply ambiguous as the utterance automatically activates all aspects of the associated lexeme for the audience. However, to use an expression requires a particular intention on the part of the speaker. Saka characterizes such use in the following way:

(u) Speaker S uses an expression *x* if and only if:

- i S produces a token of *x*, thereby ostending an open-ended number of items associated with *x*; and.

²² This feature of automatic language processing has been detailed by Fodor (1983) and Sperber and Wilson (2002).

²³ It is customary in philosophy and much of linguistics to put concepts in uppercase. I take Saka's meaning to be not that all intensions are concepts, but that the intension of a noun phrase is a concept. Sentences have intensions but it would be more contestable to say that whole sentences have concepts.

²⁴ While discussions of intensionality can be traced to Frege (1892), they have been significantly developed through work on intensional logics which is not of concern for the present project. Additionally, the use of 'intension' and 'extension' do not automatically imply that one adopts a Neo-Fregeanism about language. Finally, 'intensional' should not be confused with the philosopher's notion of 'intentional' which normally translates to 'representational'.

²⁵ See Speaks (2024, sec. 2.1.2).

- ii S intends to refer to the extension of *x* (2003, p. 190).

Thus, the use of an expression requires the speaker intending to refer to the *purported* extension of *x*.²⁶ Saka's account of mention only differs in the second condition. A speaker mentions an expression *x* if and only if "S intends to refer to something associated with *x* other than its extension" (Saka, 2003, p. 190). One way to test the intuitiveness of this theory is to see what we can predicate of a mentioned term.

[8] 'Cat' has the phonic form /kæt/.

[9] 'Cat' is a count noun.

[10] The intension of 'cat' is CAT.

[11] The meaning of 'cat' is a domesticated carnivorous mammal, classified as *Felis catus* that serves as a pet and is sometimes used for catching rats and mice (2024).

However, the following claim I take to be ill-formed:

[12] *The meaning of cat is a domesticated carnivorous mammal, classified as *Felis catus* that serves as a pet and is sometimes used for catching rats and mice (CAT, 2024).

As the lack of quotation indicators suggests, [12] is a use of 'cat.' While cats have fur and paws, they do not have meanings.²⁷

As mixed quotation involves simultaneous use and mention, accounting for it can be difficult. But it is important to note that on Saka's theory, there is nothing contradictory in what mentioning and using requires of a speaker regarding the ways they direct the audience's intentions. Both can be done simultaneously.

5.3 Slurs and quotation

5.3.1 Why quotation does not render content inert

Returning to the issue of slurs, we are in a better position to understand how slurs work in both use and mention cases. Two

²⁶ I add 'purported' to Saka's definition because, on Horn's truth-conditional account of slurs, slurring terms fail to refer. This is because the slur's content includes something like 'despicable by virtue of being a member of G' where G is the targeted group. However, as I noted in the introduction, a feature of slurs is that there is no property of G that is morally apt for derogatory force. As such, for those who take the meaning of a slur to be something like what is quoted above, slurs do not successfully pick out anything in the world, even though the people who use them most likely think they do. For non-semantic accounts of the derogatory force of slurs, where being a member of 'G' is all that is required for successful reference, the addition of 'purportedly' does no harm.

²⁷ It's worth noting that Saka does take there to be correct instances of quotation with no indicating marks. On this point, we might disagree on whether [12] is ill-formed. I take it that in written discourse, the quotation marks are at least tacit, and in spoken discourse, they can be inferred if there are no contextual indicators such as 'the word' preceding the quoted expression.

questions from the inert content criticism still need to be addressed. Why think that quotation renders content inert, and what is it for the offensiveness of slurs to project out?

Recall that the accusation against truth-conditional accounts of derogatory force is that quotation renders the content inert. We can now see why this claim fails. Any utterance of a slurring term, quoted or mentioned, will automatically activate the lexeme associated with the slur (and this is true regardless of one's theory of quotation). This would include both the intension, or content, of the term as well as the purported extension of the term. Additionally, in the case of mentioning a word, the speaker can direct the audience's attention to the intension, or content, associated with the word as well as its extension. Again, this allows us to make true claims like [11] and,

[13] The purported *customary* extension of 'cat' is {x: x a cat}.

To say that the lexeme activates the purported customary extension of 'cat' is not to say that mentions of 'cat' refer to cats. Rather, when the lexeme is activated, the hearer knows the kind of things 'cat' picks out in virtue of being a competent speaker of the language. If 'inert' means to prevent the speaker from ostending a variety of lexical features of a term, including its content, then the quotation does not render content inert.

Even if content is not rendered inert in quotation, more has to be said about how the content of an uttered slur can result in different kinds of offense and in what cases of use and quotation. I will do this in the next section.

5.3.2 Reference, derogation, and moral offense

The case where the derogatory force of an utterance is not questioned is in the standard use case of a slur. We see here that the defining feature of a use case is that the speaker intends to refer to the purported extension of a slurring term. The extension is determined by the content, some of which is derogatory. However, derogatory content is not what makes slurring utterances morally wrong.

In Section 2.1, I provided a characterization of slurs, which suggested that their content involves thinking that a group or its members are worthy of derogation by virtue of being members of a group. However, as a matter of fact, the groups that are targeted by slurs are not constituted by morally apt group properties. But, in attempting to refer with the slur, one also endorses the content of a slur. Thus, a standard use case of a slurring utterance involves both believing that a Slur accurately applies to people in the world and labeling some of those people with that slur. In other words, attempts to refer with a slur erroneously assert that there are such people who are despicable by virtue of being members of a particular group. Thus, in an attempt to refer with a slur, one commits a moral wrong: one derogates a group *erroneously and without warrant* (Gray, 2024). This is an explanation of why slur utterances can be morally offensive.

This leads to an explanation of why some cases of quotation can be morally offensive. Direct quotation, indirect quotation, and mixed quotation involve reporting what someone has said. (The exceptions to reporting are pure quotation and some cases of 'scare quotes'). Whether a speaker is targeting someone with a slur, or reporting on another person who has targeted someone with a slur, a moral wrong has been committed, and moral offense can be taken. As mentioned earlier, a common feature between Feinberg's profound offense and my moral offense is that one does not need to be directly exposed to

the offensive occurrence to take moral offense. In quoting someone else's use of a slur, a speaker makes available, via the content of the quotation, that someone has used a slur.

And what of affective offense? I described in Section 4.2.1 how understanding a slurring term can lead to affective offense, and repeated exposure to slurring terms can strengthen the association between utterances and affective offense. But this affective offense can result from and be associated with various aspects of a slurring utterance, including the slur's meaning and sound. Associative learning is nothing new and can account for some of the stranger phenomena associated with slurs, such as taking (offense) when hearing utterances phonetically similar to slurs.²⁸ In this respect, affective offense can result from both uses and mentions of slurs, including cases of pure quotation.

In the case of affective offense, we know that we have taken it because of how we feel. And while affective offense normally accompanies moral offense, it is an imperfect guide. But there is no mystery in determining whether we have taken moral offense. If we become aware of someone using a slur to derogate a person or group (either through hearing the standard usage or a report of it through quotation), a moral wrong has occurred and moral offense can be taken.

Moral offense often comes with a desire to admonish someone for their action. This desire is most naturally directed to the speaker in standard use cases; however, when a speaker is reporting a standard use case, the offense is appropriately directed at whomever used the slur to label a person or group. However, in cases of indirect quotation, such as [4], it may be unclear whether the speaker, the person being reported, or both used the slur. Clarification would be needed, and ought to be supplied, from the speaker to determine what the moral wrong is.

Another illustrative example would be a court deposition concerning hate speech. A witness might be asked to provide a verbatim report of what the defendant said. They might report:

[14] Maria: The exact words of the defendant were, "I'd be a lot happier if there were a lot less spics in our town."

In such a case, affective offense can be taken: the content of the term can be seen as repulsive and one might be emotionally effected by the mere utterance of the term. Moral offense can also be taken, but it would only make sense to direct it at the defendant.

We now have an answer to our final question concerning projectible force. Moral offense can be caused by slur mentions in the same case as slur uses: when it is clear that someone intends to refer to a person or group with a slur utterance.

Both truth-conditional and non-truth-conditional theories of slurs hold that the projectible force of slurs could project out of all or most quotational contexts. Other theorists are more skeptical of this claim (e.g., Bolinger, 2017; Cepollaro et al., 2019; Hom, 2008; Hornsby, 2001; Nunberg, 2018; Williamson, 2009). In the next section, I hope

²⁸ This is not to say it is usually permissible to utter terms that sound like slurs. Knowing that these terms sound like slurs and are likely to cause affective offense is a reason not to use them. And using them, knowing that they will most likely cause offense can be morally wrong, even if the wrongness does not result from the term uttered.

to explain this phenomenon further by examining two accounts of the linguistic mechanisms that create a slur's derogatory force and the role I take them to give to affective offense.

6 Projectible force and affective offense

There are two common responses to the criticism of inert content.

Approach 1: quotation and mentions would render derogatory content inert. Thus, whatever linguistic mechanism acts as a source of offense is not a semantic mechanism.

Approach 2: only derogatory content can be a source of offense, and quotation would render inert such derogatory content. The offense caused in cases of quotation and mention is affective offense.

It will not be possible to survey all the positions regarding approaches 1 and 2. Therefore, the cases I will look at are selected because they represent opposite ends of the linguistic spectrum regarding the derogatory content of slurs, they are well-known positions, and they have detailed explanations regarding the explanatory desideratum of projectible force. For approach 1, I will discuss Anderson and Lepore's Prohibitionist account. This approach locates the source of offense in a sociolinguistic mechanism which is a violation of a prohibition on a slurring utterance. This account almost seems custom-made to deal with the issue of projectible force. For approach 2, I will discuss Hom's combinatorial externalism, which takes derogation to be a feature of the truth-conditional content of a slur. I will then look at Rinner and Hieke's more recent defense of the truth-conditional content view, which deals specifically with the issue of slurs under quotation.

My conclusion will be that, at best, in cases of quotation that do not involve attributing a referring use of a slur to someone, offense will be limited to affective offense. Moreover, I think that the three accounts I cover would agree with this.

My criticisms of these views are not meant to be exhaustive and are limited to issues concerning their accounts of projectible force. Additionally, I believe all three views give us useful insights about how to approach the issue of projectible force. Nevertheless, not making use of the two kinds of offense I have delineated leads to some confusion.

6.1 Approach 1: a non-content-based solution

6.1.1 Prohibitionism

While intuitions do vary concerning how offensive it is to mention slurs, Anderson and Lepore (2013a) have most forcefully argued for the claim that slur mentions can cause offense. They locate the source of offense in the sociolinguistic mechanism of taboo or prohibition:

When a word is prohibited, then whoever violates its prohibition risks offending those who respect it. Presumably, prohibitions include deeply embedded occurrences of the word: embedding, we know, sometimes renders semantic properties of an expression inert, but it cannot nullify its occurrence and the prohibition is

against that. Our proposal is this: slurs are prohibited words (2013a, p. 38).

In other words, prohibition is a social phenomenon that renders certain terms unspeakable or taboo. Thus, mere utterances of terms, whether used or mentioned, can violate a prohibition and are considered taboo. It is the violation of the prohibition that is the source of offense: "they offend those for whom these prohibitions matter" (2013a, p. 46).

6.1.2 Prohibitionism and offense taken

Prohibitions can come about for moral reasons. For instance, social practices like uttering a slur could be deemed inappropriate if such practices are taken to be harmful to the targeted group. The application of slurs could also be seen as wrong for the reasons I have given above. If people with the right kinds of power decide to prohibit such words, then the continued use of such words will be seen as offensive. So, we can say that the prohibition occurs for social and moral reasons.

While the prohibition occurs for moral reasons, and using slurs can still be seen as wrong, this does not mean that offense taken at a prohibition violation is a moral offense. While the cause of the prohibition is the wrongful conduct involved in slur utterances, violating a prohibition just breaks with sociolinguistic norms—more extreme but not different in kind to breaking standards of etiquette. As such, prohibition violation is not a source of moral offense.

Violating a prohibition does cause some kind of offense. For instance, violating a prohibition on the utterance of slurring terms could cause shock, disgust, remorse, or other features associated with affective offense.

Many hold the view that slurs under quotation are less offensive than used slurs. This is the previously mentioned desideratum of synchronic-derogatory intra-variation. I've provided reasons to think that moral offense could be similar in both cases, but affective offense does seem stronger in use cases. One way of spelling out this difference is that in addition to the affective offense of prohibition violation, there is the affective offense caused by the moral wrongdoing of a slur usage. In cases of direct quotation, the affective offense could be lessened as we are not in direct contact with the slurring usage. Appealing to two different kinds of offense, derogatory content, and prohibition violations could help capture the idea that slurs under quotation can be less offensive than standard use cases of slurs.

Unfortunately, Anderson and Lepore close off this avenue. First, Anderson and Lepore avow semantic minimalism: they hold that the slur 'k*ke' and the term 'jew' have the same semantic content and only differ in that the former is prohibited. Moreover, like many slur theorists, they claim that there is only one source of offense: "Our overarching aim in this paper is to deflate all content strategies: each, no matter how it is conceived, we will argue, is irrelevant to an understanding of how slurs function and why they offend" (2013a, p. 26).

6.1.3 Lessons from prohibitionism

If one accepts the inert content criticism, it is reasonable to look at non-content-based linguistic mechanisms, such as prohibition, to explain the derogatory force in the case of mentioned slurs. But it is important to recall what we have lost before we accept what prohibitionism has given us. A content-based theory of slurs was able to explain moral and affective offense in standard use cases. However,

what Approach 1 returns to us is only an account of affective offense. While a theory of slurs ought to have an account of affective offense, it is not a replacement for the account of moral offense that was lost.

6.2 Approach 2: content-based approaches

Recall,

Approach 2: only derogatory content can be a source of offense and quotation would render inert such derogatory content. The offense caused in cases of quotation, and mention is affective offense.

I should note that anyone who takes approach two would not put it in these terms. Rather, the claim is often made by saying that slurs mentioned in utterances cannot cause (moral) offense but do result in some kind of negative affect or that they can cause offense, where this is implicitly taken to be affective offense. I will provide two examples from defenders of a truth-conditional content strategy.

6.2.1 Combinatorial externalism—Hom

Hom (2008) has developed the most well-known truth-conditional content account of a slur's derogatory force. He calls his view 'combinatorial externalism' because (1) the content of terms is partially fixed by the social practices of one's linguistic community and not just what is inside of one's head, and (2) the meaning of a slur is a complex predicate which combines the claims that members of a group ought to be subject to discriminatory practices because they have certain negative properties in virtue of being members of their group (2008, p. 431). The content of slur utterances explain why they have derogatory autonomy. If a speaker uses a slur, it is not up to that speaker to decide whether their use is derogatory. The derogatory force is a feature of the truth-conditional content and not the speaker's intention.²⁹

6.2.2 The affect of slurs under quotation

Hom (2008) lists derogatory autonomy as an explanatory desideratum, but not offensive autonomy. While Hom recognizes that people seem to take offense in cases of slur *mentions*, he does not take this to be caused by the derogatory content of the slur. Rather Hom explains 'squeamishness' in the presence of mentioned or quoted slurs. Hom argues:

Most, if not all, competent, nonracist speakers of English observe their own feelings of squeamishness that typically accompany uses of epithets. Silentism [i.e., prohibitionism] relies on this phenomenal fact to generalize to all uses of epithets, regardless of their syntactic embedding or the conversational context of their utterance. Hence, according to silentism, since squeamishness accurately tracks derogation, all uses of epithets must derogate their relevant target class. . . . The problem is that, for many, squeamishness occurs not only for epithets embedded under negation, conditional antecedents, questions, intensional and fictional contexts, but also epithets under quotation, in contexts

of appropriation, and even to mere phonological variants (for example... 'niggardly'). These observations call into question the accuracy of squeamishness as a guide to derogation, especially in the last case of semantically and etymologically distinct, phonological variants. . . . Hence because these words are so highly charged, our intuitions have limited value from the outset, and it would be hardly surprising if, at least in some cases, our intuitions were even misleading (2008, p. 435).³⁰

There are several insights to unpack here. First, I think Hom is correct that we often use squeamishness (which I take to be a version of affective offense) as an imperfect litmus test for tracking the derogatory force of a slur (what I have been calling the source of moral offense). Using the distinction between moral and affective offense we can interpret Hom as expressing the following concern about squeamishness. Since affective offense occurs in the cases of mentioned slurs as well as utterances phonetically similar to slurs, people may mistakenly take it that quoted slurs and phonetically similar utterances are also morally offensive. However, it should be clear that no moral offense ought to be taken upon hearing utterances phonetically similar to slurs. Similarly, no moral offense ought to be taken upon hearing epithets under quotation.

Thus, Hom gives us the following explanation of slurs under quotation: The derogatory force of a slur utterance is rendered inert in cases of quotation. However, there is no moral offense to be taken in such cases. Insofar as a hearer takes offense in the presence of a slurring term, this should be understood as affective offense.³¹

I take one disagreement between Hom's view and mine to concern what is occurring in cases of quotation. In a case of pure quotation, there would be agreement that there is no moral offense, as no one is labeled with the slur.

[15] Pure Quotation: "The word 'redneck' is used to refer to poor southern white rural labor workers."

If offense is to be taken, it would be something like squeamishness or affective offense. However in cases of direct, indirect, and mixed quotation where a slur is directed at a person or group, I take moral offense to be warranted. This is the case in [3] and [4]. When quotations report on referential uses of slurring terms, moral offense can be taken by virtue of the moral wrong committed by the person reported in the quotation.

6.2.3 Content as a causal source of offense—Rinner and Hieke

Recently, Rinner and Hieke (2022) have taken Approach 2 in defending a content theory in the face of prohibitionism (Anderson and Lepore, 2013a). They hold that "Slurs under quotation can cause offense" and that an adequate content theory must provide an adequate explanation of this explanatory desideratum (2022, p. 1485). They account for the offense caused by slurs under quotation in the following way:

³⁰ Bracketed text added for clarity.

³¹ I take it that for Hom 'offense' is what I am calling 'moral offense' and what I am calling 'affective offense' is, for Hom, not offense at all.

²⁹ A pragmatic view that locates the source of offense with speaker's intent is given by Kennedy (2022).

Although quotation would render the derogatory content of slurs inert, *this very content* could still play an important causal role in cognitively or emotionally affecting members of the target group. In general, *content can be causally efficacious* when an expression is only quoted. . . . the content of an expression can be psychologically efficacious even when the expression is only mentioned (2022, p. 1486).³²

As an example, Rinner and Hieke invite us to think of the following kinds of sentences:

[16] ‘Aristotle’ is the English version of a Greek name.

[17] ‘K*ke’ is a derogatory word (2022, pp. 1485–1,486).

In [16], Rinner and Hieke, claim the mentioned terms cause us to think about Aristotle. Similarly, in [17], ‘K*ke’ causes hearers “to think about the derogatory content of the . . . slur be it semantic, pragmatic, or presuppositional. . . . [quoted slurs] can still provoke negative emotions among the members of the respective target group and might in some, or even many, cases also cause offense, especially in the case of highly loaded slurs” (2022, p. 1486).³³ I take this to amount to the following. Quoting a word may cause me to think about that word, and thinking about the word causes me to think about the content of the word. If the word has derogatory content, then that derogatory content can cause ‘negative emotions and offense.’

There is an initial similarity between Saka’s view of quotation and that of Rinner and Hieke. Both acknowledge the role of lexical activation in understanding speech, whether used or mentioned and that this can evoke cognitive states in the hearer. Additionally, Rinner and Hieke discuss how they can evoke emotional states.

However, there are several unanswered questions and some concerns. We are given no account of what it means for content to be inert. I take it that if content is rendered inert, slur utterances will not have their usual derogatory force. I take it that the account of the causal influences on our cognitive and emotional states is used to explain why content can still result in offense being taken. However, even when content is not rendered inert, there are still causal relations between what a speaker says and our cognitive and emotional states. What is needed is an account of what we have in addition to causal relations in standard uses of slurs that are missing in quotation.

A point of difference between my view and that of Rinner and Hieke is the degree and kind of offense present in slurs under quotation. Rinner and Hieke note that offense can be greatly diminished when slurs are under quotation. However, they focus almost entirely on cases of pure quotation (their eight examples of slurs under quotation are all cases of pure quotation). Admittedly, cases like [17] are not morally offensive as they do not report on referential uses of a slur by direct, indirect, or mixed quotation. These would be harder cases to account for as the moral offense warranted in cases of reported slur usage is more significant and thus harder to account for if content is rendered inert.

Another point of difference is that their discussion of offense focuses almost entirely on ‘negative emotional responses’ (2022, p. 1485,1,486). This suggests that, like Anderson and Lepore and Hom, the account of

offense given for slurs under quotation is an affective account. Thus, we are left with no account of moral offense for slurs under quotation.

7 Conclusion

The three accounts covered in section 6 show that when attempting to account for the offensiveness of slurs under quotation, there is a broad appeal made to affective offense. However, neither the prohibitionist nor the truth-functional accounts we have looked at have provided an account of moral offense under quotation, which is the main kind of offense that concerns us in standard use cases. I take it that I have provided such an account of moral offense for both standard use cases and cases of quotation and mention where someone has made a referential use of a slurring term.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

DG: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

Thanks to the two reviewers at Frontiers in Communication for helpful criticisms.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

³² Italics added for emphasis.

³³ Bracketed text added for clarity.

References

- Anderson, L., and Lepore, E. (2013a). Slurring words. *Noûs* 47, 25–48. doi: 10.1111/j.1468-0068.2010.00820.x
- Anderson, L., and Lepore, E. (2013b). What did you call me? Slurs as prohibited words. *Anal. Philos.* 54, 350–363. doi: 10.1111/phib.12023
- Asher, N. (1993). Reference to abstract objects in discourse. Dordrecht: Springer Netherlands.
- Assimakopoulos, S. (2022). Ostension and the communicative function of natural language. *J. Pragmat.* 191, 46–54. doi: 10.1016/j.pragma.2022.01.021
- Ásta (2018). Categories we live by: The construction of sex, gender, race, and other social categories. New York, NY, United States of America: Oxford University Press.
- Ásta (2022). Can Conferralism account for systemic racism? *South. J. Philos.* 60, 21–36. doi: 10.1111/sjp.12461
- Bach, K. (2018). “Loaded Words,” in *Bad words: Philosophical perspectives on slurs*. ed. D. Sosa (Oxford: Oxford University Press), 60–76.
- Barrett, L. F., and Bliss-Moreau, E. (2009). Affect as a psychological primitive. *Adv. Exp. Soc. Psychol.* 41, 167–218. doi: 10.1016/S0065-2601(08)00404-8
- Beaumont, H. (2024). ‘A lack of trust’: how deepfakes and AI could rattle the US elections. *Al Jazeera*. Available at: <https://www.aljazeera.com/news/2024/6/19/a-lack-of-trust-how-deepfakes-and-ai-could-rattle-the-us-elections> (Accessed June 19, 2024)
- Bianchi, C. (2014). Slurs and appropriation: an echoic account. *J. Pragmat.* 66, 35–44. doi: 10.1016/j.pragma.2014.02.009
- Blakemore, D. (2015). Slurs and expletives: a case against a general account of expressive meaning. *Lang. Sci.* 52, 22–35. doi: 10.1016/j.langsci.2014.06.018
- Boisvert, D. R. (2008). Expressive Assertivism. *Pac. Philos. Q.* 89, 169–203. doi: 10.1111/j.1468-0114.2008.00315.x
- Bolinger, R. J. (2017). The pragmatics of slurs. *Noûs* 51, 439–462. doi: 10.1111/nous.12090
- Camp, E. (2013). Slurring perspectives. *Anal. Philos.* 54, 330–349. doi: 10.1111/phib.12022
- Cappelen, H., Lepore, E., and McKeever, M. (2023). Quotation. Available at: <https://plato.stanford.edu/archives/sum2023/entries/quotation/> (Accessed June 1, 2024).
- CAT (2024). Available at: <https://www.merriam-webster.com/dictionary/cat> (Accessed August 5, 2024).
- Cepollaro, B. (2015). In defence of a presuppositional account of slurs. *Lang. Sci.* 52, 36–45. doi: 10.1016/j.langsci.2014.11.004
- Cepollaro, B., and Stojanovic, I. (2016). Hybrid Evaluatives: in defense of a Presuppositional account. *Grazer Philos. Stud.* 93, 458–488. doi: 10.1163/18756735-09303007
- Cepollaro, B., Sulpizio, S., and Bianchi, C. (2019). How bad is it to report a slur? An empirical investigation. *J. Pragmat.* 146, 32–42. doi: 10.1016/j.pragma.2019.03.012
- Connell, F. J. (1967). Double effect. *Principle of. New Cathol. Encycl.* 4, 1020–1022.
- Croom, A. M. (2011). Slurs. *Lang. Sci.* 33, 343–358. doi: 10.1016/j.langsci.2010.11.005
- Davidson, D. (1979). “Quotation” in *Inquiries into truth and interpretation* (Oxford, New York: Clarendon Press; Oxford University Press), 79–92.
- De Brabanter, P. (2010). The semantics and pragmatics of hybrid quotations. *Lang. Linguist. Compass* 4, 107–120. doi: 10.1111/j.1749-818X.2009.00185.x
- DiFranco, R. (2017). Appropriate Slurs. *Acta Anal.* 32, 371–384. doi: 10.1007/s12136-016-0313-0, Appropriate Slurs
- DiFranco, R., and Morgan, A. (2023). No harm, still foul: on the effect-independent wrongness of slurring. *J. Am. Philos. Assoc.* 9, 471–489. doi: 10.1017/apa.2022.18
- Dummett, M. (1973). “Frege and Wittgenstein” in *Frege and other philosophers* (Oxford: Oxford University Press).
- Feinberg, J. (1987). *The moral limits of the criminal law: Offense to others*. New York: Oxford University Press.
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: The MIT Press.
- Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik, Neue Folge.* 100, 25–50.
- Glasgow, J., Haslanger, S., Jeffers, C., and Spencer, Q. (2019). What is race?: Four philosophical views., in what is race?: Four philosophical views. New York: Oxford University Press.
- Gray, D. M. (2022a). New social ontologies of race. *South. J. Philos.* 60, 1–3. doi: 10.1111/sjp.12419
- Gray, D. M. (2022b). Social ontologies of race and their development. *South. J. Philos.* 60, 4–20. doi: 10.1111/sjp.12481
- Gray, D. M. (2024). Doxastic and epistemic sources of offense for slurring terms. *Acta Anal.* doi: 10.1007/s12136-024-00609-z
- Gutzmann, D., and Gärtner, H.-M. (Eds.) (2013). Expressives and beyond: an introduction to varieties of use-conditional meaning. *BRILL*. doi: 10.1163/9789004183988_002
- Halimi, S., and Rimbart, P. (2024). Curtail free speech? ‘Oh yes we can’. *Monde Dipl.* Available at: <https://mondediplo.com/2024/06/11/freedom-of-speech> (Accessed June 19, 2024).
- Hansen, S. J., McMahon, K. L., and de Zubizaray, G. I. (2019). The neurobiology of taboo language processing: fMRI evidence during spoken word production. *Soc. Cogn. Affect. Neurosci.* 14, 271–279. doi: 10.1093/scan/nsz009
- Hardimon, M. (2022). Why social constructionists should embrace minimalism. *South. J. Philos.* 60, 37–53. doi: 10.1111/sjp.12478
- Haslanger, S. (2020). “Going on, not in the same way” in *Conceptual engineering and conceptual ethics*. eds. A. Burgess, H. Cappelen and D. Plunkett (Oxford: Oxford University Press), 230–260.
- Hom, C. (2008). The semantics of racial epithets. *J. Philos.* 105, 416–440. doi: 10.5840/jphil2008105834
- Hom, C. (2010). Pejoratives. *Philos. Compass* 5, 164–185. doi: 10.1111/j.1747-9991.2009.00274.x
- Hornsby, J. (2001). Meaning and uselessness: how to think about derogatory words. *Midwest Stud. Philos.* 25, 128–141. doi: 10.1111/1475-4975.00042
- Hunter, J., and Asher, N. (2016). Shapes of conversation and at-issue content. *Semant. Linguist. Theory* 26:1022. doi: 10.3765/salt.v26i0.3946
- Janschewitz, K. (2008). Taboo, emotionally valenced, and emotionally neutral word norms. *Behav. Res. Methods* 40, 1065–1074. doi: 10.3758/BRM.40.4.1065
- Jeshion, R. (2013). Expressivism and the offensiveness of slurs. *Philos. Perspect.* 27, 231–259. doi: 10.1111/phpe.12027
- Kennedy, R. (2022). *Nigger: The strange career of a troublesome word*. Revised Edn. New York: Pantheon.
- Liu, C. (2021). The derogatory force and the offensiveness of slurs. *Organon* 28, 626–649. doi: 10.31577/orgf.2021.28307
- Lycan, W. G. (2015). Slurs and lexical presumption. *Lang. Sci.* 52, 3–11. doi: 10.1016/j.langsci.2015.05.001
- Madan, C. R., Shafer, A. T., Chan, M., and Singhal, A. (2017). Shock and awe: distinct effects of taboo words on lexical decision and free recall. *Q. J. Exp. Psychol.* 70, 793–810. doi: 10.1080/17470218.2016.1167925
- Mallon, R. (2016). *The construction of human kinds*. Oxford: Oxford University Press. Available at: <https://academic.oup.com/book/5611>
- Mallon, R. (2022). What’s at stake in the race debate? *South. J. Philos.* 60, 54–72. doi: 10.1111/sjp.12480
- Mandelbaum, E., Ware, J., and Young, S. (2024). “The sound of slurs: bad sounds for bad words,” in *Oxford studies in experimental philosophy*. eds. J. Knobe and S. Nichols (Oxford: Oxford Academic).
- Marquis, D. B. (1991). Four versions of double effect. *J. Med. Philos.* 16, 515–544. doi: 10.1093/jmp/16.5.515
- McCready, E. (2010). Varieties of conventional implicature. *S&P* 3, 1–57. doi: 10.3765/sp.3.8
- Neufeld, E. (2019). Essentialist theory of meaning of slurs. *Philos. Impr.* 19.
- Nunberg, G. (2018). “The social life of slurs” in *New work on speech acts*. eds. D. Fogal, D. Harris and M. Moss (Oxford: Oxford University Press), 237–295.
- Popa-Wyatt, M., and Wyatt, J. L. (2018). Slurs, roles and power. *Philos. Stud.* 175, 2879–2906. doi: 10.1007/s11098-017-0986-2
- Potts, C. (2004). *The logic of conventional Implicatures*. Oxford: Oxford University Press.
- Potts, C. (2007). The expressive dimension. *Theor. Linguist.* 33, 165–198. doi: 10.1515/TL.2007.011
- Predelli, S. (2010). “From the expressive to the derogatory: on the semantic role for non-truth-conditional meaning” in *New waves in philosophy of language*. ed. S. Sawyer (London: Palgrave Macmillan UK), 164–185.
- Richard, M. (2008). *When truth gives out*. Oxford: Oxford University Press.
- Rinner, S., and Hieke, A. (2022). Slurs under quotation. *Philos. Stud.* 179, 1483–1494. doi: 10.1007/s11098-021-01715-z
- Saka, P. (1998). Quotation and the use-mention distinction. *Mind* 107, 113–135. doi: 10.1093/mind/107.425.113
- Saka, P. (2003). Quotational constructions. *Belg. J. Linguist.* 17, 187–212. doi: 10.1075/bjl.17.11sak
- Saka, P. (2007). *How to think about meaning*. Dordrecht: Springer Netherlands.
- Scanlon, T. (2010). *Moral dimensions: permissibility, meaning, blame*. Cambridge: Harvard University Press.

- Schlenker, P. (2007). Expressive presuppositions. *Theor. Linguist.* 33, 237–245. doi: 10.1515/TL.2007.017
- Searle, J. R. (2011). *Speech acts: An essay in the philosophy of language*. 34th Edn. Cambridge: Cambridge University Press.
- Sennet, A., and Copp, D. (2015). What kind of a mistake is it to use a slur? *Philos. Stud.* 172, 1079–1104. doi: 10.1007/s11098-014-0338-4
- Speaks, J. (2024). Theories of meaning. Available at: <https://plato.stanford.edu/archives/fall2024/entries/meaning/> (Accessed August 5, 2024)
- Sperber, D., and Wilson, D. (2002). Pragmatics, modularity and mind-reading. *Mind Lang.* 17, 3–23. doi: 10.1111/1468-0017.00186
- Tamás, H., Emese, M., Emese, H., Karolina, J., and Dezső, N. (2010). A tabuszavak hatása lexikális döntési helyzetben. *Psychiat Hung* 25, 525–537.
- Tirrell, L. (1999). “Derogatory terms: racism, sexism and the inferential role theory of meaning” in *Language and liberation: Feminism, philosophy and language*. eds. K. Oliver and C. Hendricks (Albany: SUNY Press), 41–79.
- Wedgwood, R. (2011). Defending double effect. *Ratio* 24, 384–401. doi: 10.1111/j.1467-9329.2011.00508.x
- White, K. K., Abrams, L., Koehler, S. M., and Collins, R. J. (2017). Lions, tigers, and bears, oh sh!T: semantics versus tabooeness in speech production. *Psychon. Bull. Rev.* 24, 489–495. doi: 10.3758/s13423-016-1084-8
- Whiting, D. (2013). It’s not what you said, It’s the way you said it: slurs and conventional implicatures. *Anal. Philos.* 54, 364–377. doi: 10.1111/phib.12024
- Williamson, T. (2009). “Reference, inference, and the semantics of pejoratives” in *The philosophy of David Kaplan*. eds. J. Almog and P. Leonardi (Oxford: Oxford University Press), 137–158.