

The Instability of Slurs

Christopher Davis and Elin McCready
University of the Ryukyus and Aoyama Gakuin University

November 19, 2018

1 Introduction

In this paper, we argue for a particular view of slurs according to which they are a semantically coherent class satisfying a set of related criteria:

- (1) An expression e is a slur if
 - (i) e semantically invokes a complex which can be used to derogate a particular group;
 - (ii) the derogation of that group functions to subordinate them within some structure of power relations supported by an actualized *flawed ideology*;
 - (iii) the group is one defined by an intrinsic property (e.g race/gender/sexuality/abledness).

The body of the paper constitutes an articulation and defense of these criteria.

In the course of spelling out our theory of slurs, we argue for the importance of distinguishing the *offense* that such terms can engender from the *derogatory* intent with which they can be used. We show that whether a term is or is not derogatory is partially independent of whether it is offensive, and that discussions about the (non-)pejorative status of a particular use of a slur should distinguish these notions. In particular, slurs can be used non-derogatorily while still being offensive.

We go on to argue that terms which are not semantically slurs can nevertheless be used to slur; we thus distinguish *slurs*, a class of lexical items satisfying the criteria in (1), from *slurring*, a species of speech act. Moreover, we argue that *derogation* is not equivalent to *slurring*, with the distinction turning on whether the derogated group stands in a position of subordination within the larger society, and how the speaker stands in this wider social structure.

The paper is structured as follows. In section 2, we present a framework within which the semantics of slur terms can be spelled out, and a discussion of how derogation is derived as a pragmatic implicature. We also discuss why derogation should be distinguished from offense in discussing the impact slurs have when uttered. In section 3 we turn our attention to less canonical examples of gender-based slurs and slurring. We argue that some of these terms constitute slurs in our technical sense, in that they invoke complexes of social norms that serve to subordinate. We also argue that gendered terms (in particular, gendered pronouns) which are in themselves not in

any sense slur terms can be used to slur. In section 4 we consider how the domain of slur terms should be defined, and a number of ‘edge cases’ that we think provide important areas for future work on slurs and slurring.

2 Unstable Invocations

2.1 Slurs: A semantic sketch

Before looking at non-derogatory uses of slurs, we begin by articulating a particular view of their semantics and pragmatics. This approach to slurs was first presented in McCready and Davis 2017, and here we summarize and build on that work. The core intuition behind this view is that slurs are *mixed expressives*, expressions which carry both at-issue and expressive content, in the sense articulated by McCready (2010) and Gutzmann (2015). Their at-issue semantic content is a simple predication of group membership, while their expressive content serves to *invoke* a complex of sociohistorical facts, attitudes, and prejudices about the group. Crucially, this expressive content does not itself entail a particular attitude on the part of the speaker toward the group; the content is in that sense non-attitudinal.

The denotation in (2) is a template for the denotation of slur terms in general, where G is the slurred group, and skull is the expressive content invoked by the slur.

$$(2) \quad \llbracket S \rrbracket = \lambda x.G(x) \blacklozenge \text{skull} : \langle e, t \rangle^a \times t^s$$

The at-issue content of a slur, according to this proposal, is a simple at-issue property, namely, the property of belonging to the slurred group G . The at-issue content is neutral, and equivalent to the neutral term specifying membership in G (if such a term exists; see section 3).

The expressive content, skull , is what makes a slur a slur. A number of analyses encode speaker-oriented content here (Potts, 2007; McCready, 2010; Gutzmann and McCready, 2015), essentially having the slur express a negative attitude on the part of the speaker toward the slurred group. Such a semantics makes non-derogatory uses difficult to account for, since speaker-oriented derogatory attitudes are generated as (not-at-issue) entailments. In our own account, the expressive content is itself not attitudinal, or at least does not entail anything about the speaker’s own attitudes. Rather, it serves to *invoke* a set of (externally-determined, contingent) historical facts, prejudices, and social attitudes about the group. By calling this term an ‘invocation’, we mean that a particular complex of historical facts and prejudices are called up and imposed on the context. Crucially, the nature of this complex is not under the control of the discourse participants, but is an external fact. The speaker, in uttering a slur, invokes this complex; what effects this has on any particular discourse participant, and indeed what aspects of this complex any given discourse participant is privy to, is in principle beyond the ken of the speaker. In McCready and Davis 2017, we used the following analogy to get at the idea:

A sorcerer stands atop a high cliff by the sea. He raises his hands and pronounces a single word; a submerged island rises above the waves, covered with cyclopean masonry and dripping sea plants which make the precise angles of the constructions

and their outlines indistinct. The sorcerer has summoned up a city of the past from beneath the sea, where it had heretofore lain invisible. The sorcerer can do so even without knowing every detail of what lies in the city, how it is arranged, or what the consequences of calling it up will be. Utterance of the summoning word is sufficient for the invocation. No one person observing the summoning can see all features of the conjured object. This includes the sorcerer himself.

Like the sorcerer, the utterer of a slur term succeeds, by that very utterance, in bringing in to view a (many times hidden or at least politely ignored) complex structure. With slurs, this structure consists of historical facts, prejudices, and social attitudes, none of which any given participant in the discourse will have full access to. In this sense, the nature of the complex is in principle unknowable, and imposing it on the context can have unforeseen and unforeseeable consequences. Formally, this invocational effect is modeled by the expressive term ☠; like other kinds of expressive content, it is *imposed* on the context (Potts, 2007). What makes a term a slur, as opposed to any other kind of expressive, is the particular *nature* of this imposed content. As we argue below, slurs invoke content with *negative implications* for a *subordinated* group that is itself defined in terms of *intrinsic* properties.

Before further discussing these definitional properties, and a variety of ‘edge cases’ that fail to satisfy one or more of these criteria, we will first examine how non-derogatory uses of slurs can be viewed within this framework. The above framework, we argue, *allows* for uses that do not commit the speaker to endorsing a particular attitude toward the slurred group, including the non-derogatory uses that are the focus of this issue. The question, of course, is what circumstances allow for this kind of non-derogatory use, and to what extent this is systematic and generalizable across different slur terms.

2.2 Whence derogation?

Before turning attention to *non*-derogatory uses of slurs, we first address the question of where their (default) derogatory force comes from in the first place. We do so within the basic formal framework sketched above. We argue that there are in fact two issues involved, one due to the conventionally encoded expressive content of the slur itself, and one due to a contextually variable inference whereby the speaker can be taken to endorse this content.

First, the invocational content of the slur, ☠. This content, as we describe above, does not entail anything about the speaker’s own attitudes. Rather, it serves to invoke and bring to contextual prominence a (possibly hidden) bundle of historical facts and social stereotypes about the group. Insofar as this invoked content is itself objectionable or offensive, an utterance of the slur itself will likely be considered objectionable or offensive. The offensiveness here derives not from anything entailed about the speaker’s attitudes; rather, the invoked content is itself offensive. In particular circumstances (for example, the utterance of an ethnic slur by someone who is not a member of the group targeted by that group), an additional layer of offense can be generated by the fact that this content was invoked by a non-group member. This second-order offense reflects questions about who should be allowed to make certain invocations, and who should not.

Above and beyond the offense associated with the invoked content itself, the use or mention of a slur gives rise to the issue of the speaker's own relationship to this content. This is partially an objective fact about the speaker's own group status (e.g. member or non-member of the slurred group, member or non-member of the privileged or oppressor group). But it is also a question of the speaker's individual attitude toward the invoked content: does the speaker endorse or accept this content? Is the speaker challenging it in some way? These questions cannot be determined purely based on external characteristics of the speaker themselves, but require inferences about the speaker's intentions and attitudes. This latter issue is therefore a pragmatic one.

From the above discussion, it seems that we can and should distinguish *offensive* uses of slurs from *derogatory* uses. The former relates to how the invoked content itself affects discourse participants, regardless of speaker intent. The latter reflects inferences about the speaker's own attitudes toward the invoked content, and by extension to members of the slurred group. Within the formal framework pursued here, offensiveness derives fairly directly from the conventionally encoded expressive content of the slur term, while derogation is derived as an implicature. It is thus possible to be non-derogatory and nevertheless offensive, a subtle distinction that we think is useful in understanding some of the debate and misunderstanding that arises in discussions surrounding the (in)appropriate use and/or mention of slurs.

A final analogy to help drive home the idea that offensiveness (inhering in the invoked content itself) and derogation (deriving from a calculation of the speaker's stance toward the invoked content and purposes for invoking it) should be distinguished. Imagine a violently misogynistic photograph, for example a pornographic depiction of rape. The object is itself offensive, at least for a large range of the population (and bracketing issues of consent). Moreover, any overt display of the object, that is, any action whereby others are made to directly perceive the pornographic object itself without consent, is offensive. We take these as facts. In the normal course of events, social norms demand that such an object not be publicly displayed, and any public display is likely to cause offense. That said, there are any number of reasons a person might have displayed the offensive material, and our reactions to the display will often differ according to what we take those reasons to be. The owner of a commercial website might display the object in order to titillate viewers and make money. Later, a prosecutor might display the object to a jury in order to argue that the material is beyond the pale and should, therefore, suffice as grounds for penalizing the owner of the website. In this situation, the jury may well be offended. But we would not take the prosecutor to be engaging in conduct equivalent or perhaps even comparable to that of the website proprietor. The difference here is clearly one of intent. The offensiveness of the object itself has in no way been ameliorated, and indeed the offense to the jurors might be in no way lessened by the court context. Nevertheless, a reasonable person could allow that the attorney's display of the object was appropriate given the context and reasons for its display. The website owner's use of the object is misogynistic and derogatory to women; the attorney's is not.

2.3 Non-derogatory uses of slurs

At this point, we have argued that the derogatory effect of slur terms is the result of an utterance-based inference, and that this effect should be distinguished from the offense occasioned by the mere utterance of a slur. As inferences that go above the (inherently offensive) conventional mean-

ing of the slur itself, derogation is thus a pragmatic issue: Who is the speaker? The audience? To what end is the slur being deployed?

In McCready and Davis 2017, we considered four paradigmatic configurations in order to begin to understand how such factors come into play. We looked at simple two-party conversations, in which the distinguishing factors are the speaker's and hearer's respective membership or non-membership in the group targeted by the slur. We schematized these contexts with the following abbreviatory conventions: *S* is a slur, *SG* is the group slurred by *S*, and *PG* is the dominant or privileged group with respect to *SG*. We argued that, by default, the four possible configurations of such parameters lead to the following default pragmatic interpretations of the slur:

1. *S* is uttered by a member of *PG* to a member of *SG*: **subordination**.
2. *S* is uttered by a member of *PG* to a member of *PG*: **complicity**.
3. *S* is uttered by a member of *SG* to a member of *SG*: **solidarity**.
4. *S* is uttered by a member of *SG* to a member of *PG*: **accusation**.

The contextual default, then, is for uses of *S* by *PG* members to be derogatory, in that the content is invoked to either subordinate an *SG* addressee (in the first kind of context) or to make a fellow *PG* member complicit in the program of subordinating members of *SG* (in the second). In both of these contexts, we get, by default, an inference that the speaker *endorses* the content invoked by *S*. This, we take it, is a large part of where the derogatory force of slurs in these contexts comes from: by uttering *S*, a member of *PG* will, by default, indicate implicit endorsement of the background prejudices and stereotypes invoked by *S*.

This puts any *PG* member in a precarious position; if they are going to utter a slur *S*, they will have to make an effort to block the default inferences associated with that utterance. Consider the following attempt at ameliorating the derogatory content of a slur *S* in a context with a *PG* speaker and *SG* addressee:

- (3) It's because he thinks of you as a *S*.

An utterance of a sentence like that in (3) by a member of *PG* to a member of *SG* could very well succeed in indicating the negative attitudes of a third party toward both the slurred group as a whole and the addressee in particular, without thereby committing the speaker to either of those attitudes. The utterance might still cause offense, insofar as the very utterance of *S*, in particular by a member of *PG*, will serve to invoke a complex of offensive background facts and bring them to contextual prominence. But despite the potential for such offense, it is still at least possible for the speaker in such a context to avoid implicating themselves as an endorser of the prejudices thus invoked.

But an utterance of something like (4) is (to put it mildly) much less likely to succeed in ameliorating the derogatory implications of the slur.

- (4) But I know you, and you definitely aren't a *S*.

Why? The speaker has explicitly excluded the addressee as a target of scorn, and thus does not derogate the addressee as an individual. But an utterance of something like (4) seems to implicate the speaker in endorsing more general attitudes about the slurred group as a whole. That is, an utterance of (4) would seem to fail because it indicates that the speaker does not reject the prejudices or properties invoked by *S* in general, but is simply excluding the addressee as a ‘special case’.

The contrast in these examples points to a subtle issue with derogation. The issue is not fundamentally one of whether the speaker endorses the invoked content with respect to a particular individual (here, the addressee), but whether they endorse the invoked content in any sense, for any member of the slurred group. To successfully ameliorate the derogatory potential of an utterance of *S*, a member of *PG* must take measures against the default inference that they endorse all or even some of the attitudes invoked by *S*. The likely failure for an utterance of (4) to do so seems to depend on how the slur is used in its predicational capacity. In reporting a third party’s attitude, as in (3), the speaker makes no commitment as to the appropriateness of predicating *S* of anyone at all. But in (4), which is not relativized to a third party, the speaker seems to endorse the possibility that *S* would be appropriately predicated of at least *some* members of the slurred group. This happens in a very indirect fashion; by making an assertion of *non*-group membership, the speaker has implicitly endorsed the appropriateness of *S* as a predicate that would appropriately describe some individuals. This, in turn, implies at least a partial endorsement of the invocational content of *S*, and is thus derogatory.

The above discussion of derogation makes the derogatory effect of slur terms look like a species of conversational implicature. But while pragmatic/contextual factors related to speaker intentions are involved in determining whether a particular utterance of a slur is derogatory, the way that the total pragmatic effect of an utterance of a slur term plays out is, in other respects, rather different than standard Gricean accounts of pragmatic enrichment. The Gricean account of conversational implicatures is fundamentally centered around speaker intentions and their recovery. An implicature happens when the speaker intends that it be communicated, and can expect that the audience will be able to recover this intention. Importantly, standard examples of conversational implicatures allow for ‘plausible deniability’; the content of an implicature can be explicitly cancelled, and when it is, no implicature survives.

Things seem to work somewhat differently with slurs. Here, it seems that their potential for damage is at least partly distinct from any recovery of speaker intention. An inappropriate utterance of a slur can give rise to offense regardless of speaker intentions, and clarification of those intentions can fail to ameliorate the offense. This is an important fact when considering non-derogatory uses of slurs, since it can easily happen that a listener (a) accurately infers the non-derogatory intent of the utterer of the slur, and (b) is nevertheless offended. Should such uses be classified as derogatory or non-derogatory?

Consider again cases of quotation, like that in (3). While we have argued that such sentences can be used non-derogatorily, the mention of a slur is still a fraught business, even in the protective confines of quotation. This is especially evident for the most volatile of slurs in (American) English: the n-word.¹ For this slur in particular, the very utterance by a non-member of the slurred

¹We make a conscious choice not to quote this particular slur, for reasons that should be made evident in the surrounding discussion.

group is regularly perceived to be an offense. For this reason, it appears that the offense triggered by the utterance (or written depiction) of a slur can go beyond the normal use/mention distinction, to the occasional consternation of slur-quoters. In such cases, we think it is again helpful to distinguish the *offense* occasioned by the utterance of the slur from the question of whether or not its use or mention is understood as derogatory.

The complications here can in part be understood in terms of the distinctions drawn above; in particular, the idea that the very utterance of a slur will result in an invocation of its expressively encoded content ☹️, and that insofar as the content invoked itself is offensive, no amount of embedding will help to diffuse the offensiveness that results. A speaker may well make their utterance under the belief that, given a lack of intent to offend or derogate, and a reasonable expectation that such an intent can be inferred, such uses or mentions of the slur term should fail to give rise to either offense or derogation. But this strategy can and does fail; even when the speaker's intent is successfully inferred, offense can still be caused. The question of whether the slur in these cases is *derogatory*, rather than simply *offensive*, then becomes a subtle question.

Despite the fact that quotation and other embedding strategies can fail to mitigate either the offensive or derogatory force of the term, and for some slurs can *systematically* fail to mitigate the offensiveness of the term, it seems clear that such strategies *can* mitigate the derogatory effect that slurs carry. The fact that they *can* do so is perhaps not surprising; what is striking is the fact that such strategies can *fail*, and that they often do in public discourse. From a theoretical point of view, the fact that these semantic strategies are not fool-proof requires an explanation. In the account of slurs sketched above, the failure of such strategies results from the invocational power the slur has. Concretely: the slur has an expressive meaning component, ☹️, which imposes onto the context a complex of historical facts, stereotypes, and prejudices. This complex is not attitudinally linked to the speaker, and thus the conventional content of the slur does not entail anything about the speaker's attitudes to the invoked complex. However, the invocation of this content is unavoidably triggered by utterance of the slur; it is in this sense that we say the content is expressive. Technically, this means that it projects through any kind of embedding it may find itself in; in dynamic terms, its content is simply imposed on the context by its very utterance. The imposed content is itself offensive, which helps explain why quotation or other speaker-distancing strategies can fail to ameliorate the offense occasioned by the utterance of the slur. Moreover, there is a kind of second-order issue about whether the speaker has the right to force this content into the context; this can help to explain why certain slurs are perceived as more offensive depending on who utters them, again regardless of the speaker's actual or inferred attitudes to the invoked content.

If the mere mention of the slur can trigger the associated invocation, and the invoked content is the thing that is offensive, then the failure of semantic strategies for ameliorating this invocation can be understood. What becomes less clear is how to account for successful cases of amelioration, and also what exactly is being ameliorated in such cases. What seems clear is the need to distinguish whether the use or mention of a term is *offensive*, and whether or not it is *derogatory*. The former question hinges, we think, on facts about the invoked content itself.² The latter question hinges on

²There are complex issues here, as we have already alluded to. In particular, take a volatile racial slur such as the n-word. Is this word equally offensive regardless of who utters it? We think not. In particular, it seems that an

what inferences can be drawn about why the speaker has invoked this content.

3 Semantic slurs and pragmatic slurring

The previous section sketched an account of the semantics and pragmatics of slurs. On this view, slurs are expressions which invoke complexes of sociohistorical facts, attitudes, and prejudices about a particular group. But this cannot be the end of the story, for if it were, the domain of slurs would fail to be properly delimited; a phrase like *white race*, for example, invokes such a complex of fact, but is not generally regarded as a slur. If anything, its use is associated with white supremacists themselves, and the invocation associated with the term would seem to have connotations of race-based identity, power, dominance, and the like. It seems to us that, for a term to be a slur, the invoked content should have *negative implications* for the group targeted by the term. On the other hand, there are terms which in and of themselves seem not to invoke content with negative implications for a group, but which can nevertheless be used to slur a particular individual. That is, a term that is itself not a slur can be used to slur in particular contexts.

In this section, we look at these issues by considering two kinds of terms referencing gender, falling on two ends of the potential-slur spectrum — the gendered slurs discussed by Ashwell (2016) on one end, and gendered pronouns as applied to various groups of people on the other — and, after providing analyses of (some of) these terms, consider how they illuminate the debate about the domain of slurring content. Our view, which we will develop further as we consider these cases, is that canonical case of *semantic* slurring happens along two dimensions: via the use of an expression that denigrates a group, and by predication of membership in that particular group. We propose that, to count semantically as a slur (at least in our sense of the term), an expression must denigrate a group which is both defined by an intrinsic property and systematically subordinated by a flawed ideology (cf. Stanley 2015 and the discussion of how ideologies relate to slurs provided by Swanson 2015). But pragmatically, expressions which are *not* semantically slurs can be used to slur. Thus, gendered pronouns can slur when addressed to trans people with the purpose of misgendering them, though they are not slurs in any semantic sense.

3.1 Norm-based gendered Slurs

Ashwell (2016) provides an interesting discussion of terms like *slut*, *bitch* and *sissy*. She argues that such terms lack neutral correlates, in contrast to what is usually claimed in the philosophical literature on slurs – namely, that slurs *must* have neutral correlates (as based on the contrast between e.g. the n-word and *black*, the latter of which is presumably nonderogatory). The key feature of

utterance of the n-word by a white speaker is itself more offensive than an utterance of the word by a black speaker. This suggests a need to distinguish the offensiveness of the content invoked by the word itself from the offensiveness of it having been invoked by a particular person. That is to say, the content ☹️ invoked by slur term *S* is itself offensive, and on top of this the speech act of uttering *S* may trigger an additional layer of offense. These are distinct offenses, we think, but further articulation of the distinction is beyond the scope of this paper. Suffice it to say that we take the invoked content to be identical regardless of speaker, and that speaker-dependent distinctions in offensiveness result from a kind of meta-norm about who is licensed to make such an invocation, or what the consequences are when members of different groups do so.

slurs like these is that they slur in virtue of norms about how certain groups of people are supposed to behave: *slut* describes a woman who has more sexual partners than she is supposed to, according to some social norm (or because some aspect of her presentation suggests that she might), *sissy* describes a man who is more sensitive than men are supposed to be according to some norm, etc. Here, the content which is used to slur is part of the at-issue meaning of the term, but is only judged slurring because of simultaneously invoked external social norms. As Ashwell notes,

Successful reclamation may therefore require changes in social norms, so that there is a neutral term available where there is not a nearby neutral demographic term. Reclamation attempts of words like *slut* themselves seek to challenge the social norms that the use of this word rests on, but this gendered slur cannot be reclaimed until those norms change. (Ashwell 2016)

These slurs, then, depend on the existence of certain social norms; on our analysis, these norms are semantically invoked by the slur. While a detailed lexical semantics for such terms would be a research project in its own right, for the purposes of this paper we can assume that, in the at-issue dimension, a term like *slut* makes a predication S which is true of a person just in case that person exhibits behavior violating sexual norm N , with N itself being invoked expressively (that is, the very mention of the term *slut* will suffice to invoke the background social norm N upon which the truth-conditional content of the term depends).³ This background norm has negative implications for the group targeted, and thus *slut* satisfies the additional criterion for slur-hood suggested above. Without such norms to invoke, it will be difficult to use such terms to slur even pragmatically; indeed, it will be hard to even make sense of them, for what does it mean to be too promiscuous with respect to feminine norm N_f or too feminine with respect to masculine norm N_m in the absence of such norms? Since there will not be a relevant norm to ground the normative statement in its at-issue semantics, the requisite support for the slurring act will not be present, as Ashwell notes. Indeed, it's unclear that the truth-conditional content will even be well-defined in absence of the relevant norm.

3.2 Pronouns and Gendered Terms

Let us consider now expressions which are pretty unarguably not semantic slurs, but which can nonetheless be used to slur. We will continue the focus of the previous section on cases where the slurring involves gender, and take as our exemplars gendered pronouns, such as *he/him* and *she/her*, and predications of gender, as in *man/woman* and *male/female*. Our main discussion will revolve around the case of pronouns, but we will turn to direct predications of gender at the end of the section.

Let us begin with pronouns. What is their semantics? There are two kinds of position extant in the literature, to our knowledge. The two have in common that pronouns denote a free variable (or discourse referent, depending on the precise background theory). They differ in how they

³As a gendered slur, *slut* canonically targets women; the same goes for *sissy*, which canonically targets men. These restrictions might be handled via presuppositions or via the invoked norms themselves; the details are of interest but beyond the scope of the present discussion.

handle the gender specification of pronouns.⁴ The first theory takes this specification to be pre-suppositional (e.g. Sauerland 2008; Sudo 2012); this means that either the background provides information about the gender of the referent associated with the free variable or that information is accommodated (Beaver and Zeevat, 2007). The second theory takes the gender specification to be conventionally implicated (or expressive, though these won't differ much for the present case; we thus lump the two together under the cover term CIE, following McCready 2010). This means that the gender information can always be new information, though, like other instances of CIE content, this repetition need not induce infelicity due to redundancy (Potts, 2007; McCready, 2018). Both these views take pronominal gender marking to be not-at-issue content.

Here is an analysis of pronouns which takes them to be CIE content, roughly following Gutzmann and McCready (2014).

$$(5) \quad \llbracket he \rrbracket = x \blacklozenge man(x) : e^a \times t^s$$

$$(6) \quad \llbracket she \rrbracket = x \blacklozenge woman(x) : e^a \times t^s$$

We do not think there is anything inherently slurring about this content, as already pointed out. Women are subordinated in patriarchal society, and therefore use of otherwise neutral gendered terms can have negative effects, e.g. effects relating to epistemic injustice (Fricker, 2007) such as the downgrading of epistemic authority (McCready and Winterstein, 2017).⁵ But we believe that these negative aspects are not conventionally associated with the terms as such; technically speaking, as neutral terms for gender, these pronouns do not themselves invoke such content.⁶

Still, this doesn't mean that these expressions can't be used with a slurring effect. Consider the case of misgendering (see Conrod 2017 for a discussion within the linguistic literature): use of an improperly gendered pronoun to refer to a person, for instance using *she* to refer to a trans man, or *he* to refer to a trans woman; or, alternatively, using the improper category term to describe such a person, as with *woman* for a trans man, or *man* for a trans woman. Misgendering is sometimes done in error, as when a person fails to recognize the gender of the person they are speaking to (which is not, of course, limited to interactions with transgender people, as cis people are often misgendered as well). But it can also be done deliberately. This is the case we want to discuss here.

Before proceeding, a terminological note. The nature of gender and how it should be analyzed is a fraught question which we are not in a position to address in this paper (e.g. George 2016; George and Briggs 2016). We are even less in a position to offer an analysis of gender transition,

⁴The discussion here concerns gender features in languages like English, where gender is essentially notional, not grammatical, in the sense that gender indicates the gender of the individual picked out by the free variable, rather than simply being a kind of agreement (McConnell-Ginet, 2013). The situation is complicated for languages that combine notional and grammatical gender, and in which gender plays a larger role in the grammatical system.

⁵We put aside feminist concerns about the very concept of gender roles, such as those raised by Haslanger (2000), for the purposes of the present essay.

⁶We can imagine a language with, for example, both a neutral and a disparaging feminine pronoun. The former would have a neutral semantics like the one here, while the latter would explicitly invoke a set of historical facts and social attitudes whereby women have been subordinated. We are not aware of such a language. But we think something like this distinguishes the term *woman* from the term *bitch* on one of its uses, where the latter is used a generic term to refer to women in general.

which is still more complex and requires a very careful hand, even just in the terminology used to start the discussion. We therefore have chosen to adopt (as far as we know) new terminology in order to avoid presuppositions and dogwhistles that certain terms carry.⁷ Acknowledging that some may be unhappy with these choices of usage, we will use the terms ‘assigned gender’ and ‘presenting gender’ to refer to the gender which people were assigned at birth and the gender they currently live in, respectively (in a sense meant to include people who identify as having no gender or multiple genders). Trans people are then those whose assigned gender and presenting gender differ, while cis people are those whose assigned and presenting genders match. We have tried to pick terms which are as neutral as possible. Our goal here is to avoid issues of agency and choice in gender, questions about gender essentialism, and the distinction (or lack thereof) between gender and sex. Terminology in this area is difficult and constantly under revision, and we recognize that some may be unhappy with our decisions.⁸

Now: what happens when a person is referred to by a pronoun which references their assigned gender rather than their presenting gender? Trans people have explicitly decided not to live as their assigned gender. This means that using a term which predicates that gender of them rejects the identity under which they live, and also rejects their agency in deciding who they are in life. For an analogy, imagine a person who works in academia and has chosen to be an academic; it is offensive to ignore this person’s expertise and treat them as if they lack knowledge of their area of research, as in the many cases in which female academics are ‘mansplained’ to, talked over, or otherwise have their professional identity and contributions ignored or minimized. The case of misgendering is similar (aside from the issue of gender identity as a choice) but operates at a far more fundamental level of personal identity, as one’s gender identity informs most domains of one’s social existence, and one’s relationship with one’s body.

When someone misgenders by mistake, it might be hurtful, but it is an error. Since predications of gender are not inherently slurring, there is no pejoration here; using the terminology introduced earlier, we might say that it is potentially *offensive*, but not *derogatory*. Deliberate misgendering is different. When misgendering is deliberate, the speaker *intends* to indicate his rejection of the misgendered individual’s gender identity; this is a denial of an important aspect of that person’s self-conception, and indeed a rejection of their self-determination. Often this kind of utterance is made as a part of a performance of a particular ideology about gender and how it works: for instance, it might be part of a generalized rejection of trans identities. This rejection and consequent denial of the misgendered individual’s lived reality and agency is derogatory, because it is not only offensive but works to subordinate the misgendered individual within a trans-exclusionary ideology. If this is right, we must conclude that not only are there non-derogatory uses of slurs, but also derogatory uses of non-slurs. In the case of misgendering, a semantically neutral term can be used to *pragmatically* invoke a particular (trans-exclusionary) ideology, which has negative implications for the target of the term; the slurring effect is parallel to that generated by semantic slurs, but is derived through purely pragmatic mechanisms.⁹

⁷The degree to which the terminology itself is a problem is rarely acknowledged within discussions of the nature of gender and gender transition, within academic work and outside it.

⁸Thanks to Lauren Ackerman for discussion here.

⁹We want to dispose of one possible objection. It is sometimes claimed that there is a distinction between ‘biological sex’ and ‘gender’ that supports the use of pronouns in the way that we are categorizing as misgendering. For

4 Delimiting the Domain of Slurs

The preceding sections have touched on a wide range of terms with what we might call ‘derogatory potential’. We considered how the derogatory effect of such terms arises, and circumstances that facilitate or mitigate this derogatory potential. We do not, however, want to equate slurs with any and all terms with a derogatory function; as we saw, some terms that can be used to slur are canonically slurs and some are not. We think it is useful to think of slurs as a subset of potentially derogatory terms, namely, those satisfying the criteria described at the beginning of the paper, repeated as (7). Any term satisfying all of these three criteria is uncontroversially a slur; but terms not satisfying one or another criterion may be more up for debate.

- (7) An expression *e* is a slur if
- (i) *e* semantically invokes a complex which can be used to derogate a particular group;
 - (ii) the derogation of that group functions to subordinate them within some structure of power relations supported by an actualized *flawed ideology*;
 - (iii) the group is one defined by an intrinsic property (e.g race/gender/sexuality/abledness).

In the above sections, we began with a simple formal theory of slurs, according to which they are mixed expressives predicating membership in some group *G* and simultaneously invoking a host of historical facts and social attitudes about *G*. This theory corresponds to criterion (i) in (7). While we think this provides a solid foundation for thinking about both the semantics and pragmatics of slur terms, we do not think this is a *sufficient* condition for slur-hood of a term, or at least not for status as a prototypical slur.

Consider the terms ‘honky’ and ‘cracker’ in American English. We suspect that the basic formal approach outlined here will provide a good basis for the denotation of these terms; they simultaneously predicate membership in a particular racial group (‘white’) and invoke historical facts or attitudes about that group. As discussed above, we think it is moot what exactly the invoked content consists of, and that the invoked content is unstable across contexts and audiences. Nevertheless, it is obvious that the invoked content is negative. Are these terms then slurs?

The answer is of course a matter of definition. These terms certainly satisfy criteria (i) and (iii) of (7), and under some definitions this may suffice to make such terms count as slurs. Schlenker (2007), for example, explicitly takes *honky* to be a slur. Whether we call them slurs or not, we want to distinguish such terms from prototypical slurs like the n-word. Terms like *honky* are formally parallel to ethnic slurs like the n-word, but crucially differ in terms of power dynamics. In essence,

example, a person who deliberately uses a pronoun inconsistent with a trans person’s presenting gender might say that, for them, the gender specification of pronouns references ‘biological sex’ rather than ‘gender’, so their usage is not incorrect. Here ‘biological sex’ is taken to be identical to assigned gender, and a version of an essentialist view is deployed, so a change in presenting gender is supposed to be irrelevant to specification of biological sex. The empirical question is then whether this way of understanding pronouns is a natural one or simply a view developed in service of a particular (trans-exclusionary) ideology. As far as we know, the predication of gender in the not-at-issue content of pronouns is no more or less tied to assigned or presenting gender than at-issue predications of *man* or *woman*. An appeal to assigned gender in the semantics of pronouns is therefore no more convincing for misgendering here than in ordinary predication; an unwillingness to recognize a person’s presenting gender is not innocent in either case, and is slurring in its rejection of the trans individual’s agency.

white people (however defined) are and have long been in a position of power in the United States and elsewhere, and any use of a putative ‘slur’ that targets them cannot fail to be made from a position of subordination. By contrast, other racialized groups have, in the American context, experienced a host of historical indignities and negative societal stereotypes, and have done so from a position of relative powerlessness. Thus, while all of these terms involve flawed ideologies in the sense of Stanley (2015), in that both involve systematic derogation of a particular group which is not necessarily deserved by all of its members, only one involves a flawed ideology which is implemented within the broad society, in which black people are systematically deprived of power and socially subordinated. The term ‘slur’ is prototypically used for terms that target groups which stand in such a position of subordination, a theoretical move which results in criterion (ii) of (7).

Terms of contempt that target particular political orientations have been contentious recently, both within philosophy and in more general contexts. Examples from recent American political discourse include ‘rethuglican’ (targeting Republican partisans) or ‘libtard’ (targeting those with a (socially) liberal political stance). We think that, here again, the basic formal apparatus introduced earlier is a good starting point for the analysis of these terms; in both cases they predicate membership in a group (Republican and liberal, respectively) and simultaneously invoke disparaging attitudes about that group. But are these slurs?

Canonically, slurs denigrate individuals or groups on the basis of their inherent or indissociable properties: race, national origin, sexual orientation, and abledness.¹⁰ Insofar as they target a group whose identity is non-intrinsic, we suggest the answer is no. But there are complications. Both of these words are blends whose expressive content relies on a combination of the neutral term with a pre-existing expressive term, ‘thug’ and ‘(re)tard’, respectively. The latter is unquestionably a slur, which predicates group membership (mentally disabled) and invokes a complex with negative implications for that group. While the term ‘libtard’ is not a slur in terms of how it invokes content about liberals *per se*, we think it achieves its effect by ‘bootstrapping’ invoked content targeting another group, and that *this* invocation *does* constitute a slur. A similar analysis applies to the term ‘wigger’, which is used to describe white people perceived as embracing or displaying characteristics stereotypically associated with black culture. The slur relies on a blend between ‘white’ and the n-word, and achieves its effect by invoking the n-word itself, and thereby invoking the content associated with that term. As such, it is again a slur, despite the fact that the people it targets for predication are white.

Turning again to terms that target individuals on the basis of political and social beliefs, we can consider the term ‘TERF’. ‘TERF’ abbreviates ‘trans exclusionary radical feminist’ and was coined as a ‘deliberately technically neutral description of an activist grouping’ by a feminist in order to distinguish people with this belief set from other kinds of radical feminists.¹¹ The term has subsequently become viewed as a slur by people subscribing to the ideology described by the

¹⁰Slurs targeting religion might be a counterexample, but we are unaware of expressions that are both clearly slurring and clearly target religion rather than race (as it seems to us that expressions denigrating Jews and Muslims are somewhat indeterminate between race and religion). *Papist* might be one example, however. It is also possible that religious groupings have a social status like that of inalienable properties, although in principle they are simply associations based on shared beliefs.

¹¹https://www.transadvocate.com/terf-what-it-means-and-where-it-came-from_n_13066.htm, retrieved November 9, 2018.

acronym, and certainly it has been used with derogatory force by some people. But is it technically a slur in our sense? It certainly satisfies condition (i); it invokes a complex of historical fact, the hostility toward trans people expressed by (some members of) this group, which can be used by trans-inclusive people to express anger or derogation toward this kind of trans-exclusionary person. It clearly does not satisfy condition (iii), because holding this ideology is completely optional and in no sense an inherent property. If this condition correctly delimits the domain of slurs, then ‘TERF’ cannot be a slur after all for it too is not an inherent property, but an ideological one.

What about condition (ii)? Is the group described by ‘TERF’ subordinated by a flawed ideology? Likely, those trans people who are targeted by trans-exclusionary people (or who feel themselves to be so targeted due to, for example, slurring uses of gendered terms such as incorrect pronouns or sex categories) would answer with an emphatic no: rather, the trans people themselves are being subordinated by a group that has issues with how they self-conceptualize based on a trans-exclusionary ideology, and so the use of ‘TERF’ is an instance of ‘punching up’ and doesn’t count as slurring in the sense we have suggested.¹² However, some with the relevant ideology find the opposite to hold; because they deny the presenting gender of trans women, they conceptualize the relevant subordination relation to be that holding between women and men, where it is certainly the women who are subordinated by a flawed ideology. We want to suggest that much of the debate that surrounds questions about what counts as a slur involves this kind of instability about what kind of subordination is salient. The question of whether a particular term is or is not a slur, full stop, is therefore very difficult to conclusively answer given the availability of multiple subordinating structures shown to exist by intersectional feminism (hooks, 1984). This, we suggest, is one reason for the intractability of debates around these issues.

We also considered terms that, while in their normal use are not in any sense slur terms, have the potential to be *used* with a derogatory effect that is parallel to that of conventional slur terms. We introduced gendered pronouns as examples of such term. It is also possible to treat names in a similar way (though we will not provide a formal treatment here, the seed is already found in Gutzmann and McCready 2014): when an individual stops using a previous name and indicates a desire to be known by a different one, not using the new name represents a denial of that individual’s life choices and, thus, their agency. Examples would include refusing to use the new last name of an individual who changes their name as a result of marriage, or deadnaming a trans person. The core distinction to standard cases of slurring is that, with slurs, the invoked content is *inherently offensive*. This is why they are so difficult to use or even mention without giving offense, and why it can be very difficult to mitigate their derogatory potential, which stems directly from the offensiveness of the invoked content itself. In the case of misgendering or deadnaming, there is nothing about the invoked content *per se* that is offensive; it is neutral. The derogatory potential of these terms comes from the (intentional) misapplication of this invoked content.

¹²Note that some subscribing to this ideology would say they are not trans-exclusionary, because (for example) they, as feminists, support trans men because of their assigned gender. We take it that this purported support is in fact a denial of the agency and self-determined identity of trans men, and that it therefore doesn’t count as genuine support, perhaps because it excludes the very category of ‘trans man’.

References

- Ashwell, Lauren. 2016. Gendered slurs. *Social Theory and Practice* 42(2):228–239.
- Beaver, David and Henk Zeevat. 2007. Accommodation. In G. Ramchand and C. Reiss, eds., *Oxford Handbook of Linguistic Interfaces*. Oxford.
- Conrod, Kirby. 2017. *Pronouns and Misgendering*. Master's thesis, University of Washington.
- Fricke, Miranda. 2007. *Epistemic Injustice*. Oxford University Press.
- George, B.R. 2016. What even is gender? Manuscript, CMU.
- George, B.R. and R.A. Briggs. 2016. Science fiction double feature: Trans liberation on twin earth. Manuscript, CMU and Stanford University.
- Gutzmann, Daniel. 2015. *Use-Conditional Meaning*. Oxford University Press.
- Gutzmann, Daniel and E. McCready. 2014. Using descriptions. In C. Piñón, ed., *Empirical Issues in Syntax and Semantics 10*, pages 55–72. CNRS.
- Gutzmann, Daniel and E. McCready. 2015. Quantification with pejoratives. In R. Finkbeiner, J. Meibauer, and H. Wiese, eds., *Pejoration*, Linguistics Today, pages 75–102. Amsterdam, Philadelphia: John Benjamins.
- Haslanger, Sally. 2000. Gender and race: (what) are they? (what) do we want them to be? *Noûs* 34(1):31–55.
- hooks, bell. 1984. *Feminist Theory: From Margin to Center*. Routledge.
- McConnell-Ginet, Sally. 2013. Gender and its relation to sex: the myth of 'natural' gender. In G. Corbett, ed., *The Expression of Gender*. Mouton de Gruyter.
- McCready, E. 2010. Varieties of conventional implicature. *Semantics and Pragmatics* 3:1–57.
- McCready, Elin. 2018. Honorification and social meaning. To appear, Oxford University Press.
- McCready, E. and C. Davis. 2017. An invocational theory of slurs. In *Proceedings of LENLS14*.
- McCready, E. and Grégoire Winterstein. 2017. Testing epistemic injustice. *Investigationes Linguisticae* XXXVII. to appear.
- Potts, Christopher. 2007. The expressive dimension. *Theoretical Linguistics* 33:165–198.
- Sauerland, Uli. 2008. On the semantic markedness of phi-features. In D. Harbour, D. Adger, and S. Béjar, eds., *Phi Theory*. Oxford: Oxford University Press.
- Schlenker, Philippe. 2007. Expressive presuppositions. *Theoretical Linguistics* 33(2):237–245.

Stanley, Jason. 2015. *How Propaganda Works*. Princeton University Press.

Sudo, Yasutada. 2012. *On the Semantics of Phi Features on Pronouns*. Ph.D. thesis, MIT.

Swanson, Eric. 2015. Slurs and ideologies. Ms, U. of Michigan; to appear in *Ideologies*.