



No Harm, Still Foul: On the Effect-Independent Wrongness of Slurring

ABSTRACT: *Intuitively, a speaker who uses slurs to refer to people is doing something morally objectionable even if no one is measurably affected by their speech. Perhaps they are only talking to themselves, or they are speaking with bigots who are already as vicious as they can be. This paper distinguishes between slurring as an expressive act and slurring as the act of causing a psychological effect. It then develops an expression-focused ethical account in order to explain the intuition that slurring involves an effect-independent moral wrong. The core idea is that the act of expressing a morally defective attitude is itself pro tanto morally objectionable. Unlike theories that focus only on problematic effects, this view is able to shift the moral burden of proof away from victims of slurring acts and onto speakers. It also offers moral guidance with respect to metalinguistic and pedagogical utterances of slurs.*

KEYWORDS: slurs, pejoratives, inner speech, expression, derogatory attitudes

Introduction

In this article we offer an account of the ethics of slurring. Though many theorists have had much to say about which semantic or pragmatic theory best captures the distinctive features of slurs, much less has been said about what is wrong with using them. Instead, past theories have tended to focus on reactions to the use of slur words, purporting to explain how their use brings about negative psychological or behavioral effects of various kinds, including lowering the audience's estimation of the target, distorting the interlocutor's perceptions, inciting violence, eliciting shock, or causing offense (Tirrell 2017; Sennet and Copp 2015; Anderson and Lepore 2013; Feinberg 1985). Moreover, extant theories in the slurring literature have not addressed the moral implications of slurring behaviors involving nonlinguistic symbols, like the swastika, and impressions, like blackface. Rather than giving a moral assessment of slurring, theorists have been interested primarily in identifying the mechanisms by which slurs bring about their characteristic negative effects. Nevertheless, given the literature's focus on effects, one might expect the moral assessment of slurring to be simple and straightforward: slurring is wrong in virtue of its negative impact on speakers and listeners. We argue that an account along these lines would be inadequate.

Our argument proceeds as follows. The fundamental wrong-making feature of slurring resides in the expressive act the speaker performs, regardless of whether



that act is likely to result in morally problematic effects. To motivate this view, we offer two kinds of examples of slurring utterances that seem morally objectionable despite the fact that they are unlikely to produce negative effects: (1) *solidarity slurring*, wherein bigoted speakers use slurs in conversation with one another; and (2) *solitary slurring*, wherein a bigoted speaker uses a slur in an episode of private or inner speech, and so no audience is present. These acts still seem morally problematic even if we stipulate that they can make no relevant difference (e.g., if the speaker and audience are already as prejudiced as they can be, other activities are already sufficient for sustaining their attitudes, etc.). To explain this moral intuition, we first develop an account of expression according to which expressive acts involve both manifesting and endorsing an attitude. We then use this account of expression to construct an intuitively plausible moral principle, the *inheritance principle*, which captures the effect-independent wrongness of the examples discussed above. According to inheritance, if a particular derogatory attitude is morally defective, then it is wrong to express that attitude as well. Whenever a slurring act involves expressing a morally objectionable derogatory attitude, inheritance will imply that it is morally wrong.

Though the inheritance principle is intuitively plausible, some independent justification is warranted. In particular, the inheritance principle depends on the thesis that derogatory attitudes themselves can be morally objectionable even when unlikely to result in negative effects. Drawing on insights from Macalester Bell, we explain how harboring a derogatory attitude can involve failing in one's moral duty to be affectively open, that is, in the duty to avoid reacting to others with hostility or ill will (unless one has a good reason to do so) and to be open to appreciating the good qualities of others. One virtue of the inheritance principle is that it shifts the moral burden of proof from individuals and groups targeted by slurring acts (to prove that they have been affected negatively) and onto slurring speakers (to prove that they are not expressing objectionable attitudes). To further demonstrate the power of our account, we discuss how it can provide action guidance for metalinguistic utterances of slurs in academic and pedagogical contexts.

We begin in section 1 by discussing the examples of solidarity and solitary slurring that would create problems for an effects-based account of the ethics of slurring. In section 2 we develop our account of expression and use it to formulate the inheritance principle. In section 3 we describe the independent justifications for our account, and we conclude our paper in section 4 with a discussion of how some instances of expressing a derogatory attitude could be morally valuable.

1. What is Slurring, and Why Are Negative Effects not the Only Thing Wrong With It?

In this section we argue that an adequate account of the ethics of slurring must explain not only what is wrong with causing negative effects, but also why the act of slurring itself seems morally objectionable (independent of any effects). The first step is to clearly articulate the phenomena under investigation. Theorists standardly distinguish expressive acts that disparage a target *T* on the basis of characteristics or behavior specific to *T* (e.g., calling someone a 'jerk') from acts

that disparage *T* on the basis of *T*'s membership in a particular group, such as their race, ethnicity, gender, sexual orientation, etc. (Saka 2007: 148; see also Hay 2013). We take slurring acts to be expressive acts of the latter variety.

Importantly, slurring acts may, but need not, employ a conventional slur word. One may perform a slurring act targeting East Asians, for instance, simply by mimicking a stereotypical accent. Hence, an account of slurring also needs to distinguish the use of 'slur' as a verb that refers to an expressive act that disparages an individual or group on the basis of (alleged) group characteristics, and the use of 'slur' as a noun that refers to a conventional vehicle designed for use in slurring acts (Anderson and Lepore 2013: 352). The most paradigmatic cases of slurring involve speakers making assertions in which conventional slur words are used to target individuals or groups. However, slurring acts can also involve using slur words in nonassertive speech acts (like questions), embedding slur words (e.g., in negations or conditionals), or even merely mentioning slur words (Hom 2012). Other slurring acts, such as the mimicking case above, may not even feature conventional slur words at all (for more examples see Jeshion 2013; Kandil 2020; DiFranco 2017a).

Like slur words, slurring acts come in a dizzying variety of forms. For our purposes here, we will define slurring as any act such that a member of the target group could reasonably interpret that act as having disparaged them in virtue of their group membership. This definition follows a long-standing tradition in legal theory according to which determining whether behavior constitutes sexual harassment requires examining it from the perspective of a reasonable victim in the plaintiff's shoes. Writing in support of this tradition, Jones and Wade (2020) argue persuasively that asking whether a reasonable person would regard an instance of behavior as harassment may not reveal the true nature of the behavior. Instead, one should ask whether someone with the relevant intersecting identities of the victim (e.g., a young black woman) could reasonably interpret the behavior in question as harassment. Similarly, we suggest that members of a target group may be in a unique position, in virtue of their identities and experiences, to recognize the disparaging qualities of an expressive act. Adverting to a more general 'reasonable person' standard may cause us to miss these features.

Also bear in mind the distinction between a slurring utterance's ostensible target and the utterance's conventional target group (Jay 2009: 83; see also DiFranco 2017b: 380). The ostensible targets of a slur *s* are the individual(s) referred to with a slurring utterance of *s* on a particular occasion of use, while the conventional target group is the group to which *s* is conventionally applied. The ostensible target of a slurring utterance is often a member of the target group, though they need not be. For instance, a bigot who misidentifies a cisgender heterosexual woman as a member of the LGBTQ+ community could target her with a conventional slur for members of that community. Henceforth, we will use the term 'target' to refer to the ostensible target of an utterance, and 'target group' to refer to a slur's conventional target group.

One last clarification is needed before shifting to ethical analysis of slurring behavior as the term 'expression' is often used in different ways. To home in on

the precise concept we have in mind, consider the way Dorit Bar-On formulates Sellars's three senses of 'expression':

EXP₁ the action sense: a person expresses a state of hers by intentionally doing something.

EXP₂ the causal sense: an utterance or piece of behavior expresses an underlying state by being the culmination of a causal process beginning with that state.

EXP₃ the semantic sense: e.g., a sentence expresses an abstract proposition, thought, or judgment by being a (conventional) representation of it (Bar-On 2004: 216).

When describing the behavior of a slurring speaker, it is important to differentiate these expressive dimensions. If, in the course of slurring, the speaker utters a sentence that semantically expresses a proposition, then he counts as having expressed₃ a proposition. If the speaker's utterance is also a direct causal result of one of his underlying states, say anger or fear, then his utterance is a causal expression₂ of that state. Because this is primarily a moral inquiry, our focus will be on intentional actions. Hence, when we speak of slurring as an 'expressive act' in what follows, we are referring to the sense in which a speaker may, by intentionally doing something, *action-express* (EXP₁) a derogatory attitude. In many cases such acts may also at the same time constitute expression in the EXP₂ and/or EXP₃ senses. However, the locus of moral evaluation resides in the intentional act itself. As we will see below, cases of EXP₁ can be morally objectionable even when no EXP₂ or EXP₃ is present. And instances of EXP₂ and EXP₃ that do not involve EXP₁ (e.g., a Freudian slip, a confession to a therapist, etc.) appear unobjectionable.

With these theoretical tools on the table, let us now turn to our claim that an adequate account of the ethics of slurring must explain why wrongness persists even when no problematic effects are produced. Though this intuition has been expressed before (e.g., in Ashwell 2016: 228), it has not been defended. And it may seem at first like a surprising desideratum. After all, in some of the few instances where theorists have attempted moral analyses of slurring behavior they have tended to focus on negative effects. For example, Steven Gimbel claims that racist humor becomes morally objectionable only when its 'roughness penetrates into the real world in a way that causes a certain sort of harm' (Gimbel 2018: 150). Christy Mag Uidhir argues that racial caricature in film (e.g., blackface) involves promulgating misinformation about race and is wrong in proportion to the harm that is likely to result from misleading audiences about morally relevant actual-world states of affairs (Mag Uidhir 2013: 63–65). For instance, cultural and historical facts about Indigenous groups in the United States seem morally relevant, and it is easy to see how misinformation promulgated by Hollywood's depictions of Indigenous people (especially in westerns made during the mid-twentieth century) may have caused harm. However, as the following harmless examples should make clear, any ethical account of slurring that relies solely on negative effects is incomplete.

Let us call cases like (A) instances of solidarity slurring:

- (A) Imagine two bigoted speakers who use a slur for a group *G* in conversation with one another while no *G*s (or allies) are present. Since *ex hypothesi* both hearers are bigots, neither is offended. Suppose further that the speakers' prejudice toward *G*s is already extreme, such that individual slurring acts are causally inert—they do not increase the likelihood that the speakers will do anything objectionable to *G*s and do not produce any measurable change in the degree of their prejudice. Moreover, each speaker recognizes the strength and stability of the other's attitudes toward group *G*, and therefore their utterances are not 'morally risky' from an effects-based point of view.

Because slurring does not require the presence of an audience, we may generate examples from solitary speech as well. Even when a speaker does not intend to address an audience apart from herself, she can still intentionally act in such a way as to give voice (or purport to give voice) to an inner state. Private speech (speaking aloud with no audience present) and inner speech (speaking silently) are thus potential vehicles for expressive acts (Morgan 2019). Let us call cases like (B) instances of solitary slurring:

- (B) Imagine a speaker who uses a slur for a group *G* during episodes of private speech (audible speech, but no audience present) and inner (silent) speech. This speaker is an extreme bigot, so his individual uses of slurs are causally inert in the sense that they are unnecessary to incite him to do anything objectionable, and they do not produce any measurable change in him (he is already as bigoted as he could be). The speaker is aware of the strength and stability of his perspective toward members of group *G*, and therefore his slurring behavior is not 'morally risky' from an effects-based point of view.

In cases like (A) and (B) utterances of slurs do not offend or demoralize target group members (or their allies). Nevertheless, their behavior still seems morally objectionable.

One way in which an effects-based ethical account might try to explain the wrongness of solitary slurring is by appealing to the effects speakers can have *on themselves*—for example, by turning themselves into a racist or by making themselves more racist. However, such a response fails to appreciate the full scope of the challenge posed by examples (A) and (B). In these cases, the slurring behavior is stipulated to make little to no difference to the speakers themselves because they are already as bigoted as they can be.

Alternatively, an effects-based theorist could argue that solidarity and solitary slurring are wrong in cases like (A) and (B) because they function to sustain speakers' maximally bigoted perspectives. However, in many cases the speakers'

own expressive acts may be unnecessary for maintaining their perspective. For example, imagine that in (A) and (B) it was the speakers' consumption of social media that maintained their bigotry. In that case, their perspectives would have been reinforced whether or not they themselves had ever performed slurring acts—their speech on its own made no causal difference. Moreover, even in cases where an isolated slurring act does make a minor difference to a speaker's attitudes and dispositions, the effect may be incredibly slight—the contribution a token slurring utterance makes in marginally reinforcing or minimally intensifying prejudiced attitudes does not fully explain the extent to which it is wrong for the speaker to slur.

A somewhat different style of effects-based approach may appeal to rule consequentialism. Instead of assessing individual actions, such a view would determine what is right by appealing to rules that have good consequences when widely followed. Presumably, a general moral norm against slurring is likely to have better consequences than allowing slurring in some contexts. On this account, the grounds for condemning private and inner speech uses of slurs is that they violate a socially beneficial rule. Thus, the rule consequentialist concludes, objectionable effects are the crucial normative consideration after all.

To illustrate the problems with rule consequentialist approaches, consider the following objection inspired by Anderson (2015) and Richter (1986). Imagine a situation where all racists are banished to an isolated private island where they are unable to harm anyone. Following a rule that only allows telling racist jokes as long as one is an inhabitant of this island may actually have a net benefit because racists find such jokes amusing and telling them does no harm. Telling racist jokes nevertheless seems morally objectionable and an effects-based rule consequentialist account cannot explain why.

Moreover, such an account will be susceptible to standard criticisms of rule consequentialism in general (Hooker 2015). Consider a rule that says slurring is permissible only in solidarity and solitary contexts. This is plausibly the kind of rule currently used by many prejudiced speakers who avoid slurring when in the presence of members (or allies) of the target group. Intuitively, an account that can judge slurring acts to always be *pro tanto* morally wrong seems preferable. The rule consequentialist cannot explain why.

At the end of the day, effects-based theorists may want to bite the bullet, insisting that the absence of negative effects entails that no wrong is incurred. To see the unintuitive implications of this view, observe that solitary utterances of slurs can make a difference only to the attitudes and dispositions of nonprejudiced or mildly prejudiced speakers. Therefore, effects-based theorists must admit that, *ceteris paribus*, solitary slurring by nonracists is morally reprehensible (because it can have measurable negative effects), while solitary slurring by extreme racists is *not* wrong (because it cannot make them any worse). On such views, solidarity slurring amongst racists would not be wrong either. An account that lets the extremely prejudiced in particular off the hook for slurring because they are already prejudiced is a very strange view indeed!

2. An Expression-Centered Account of the Wrongness of Slurring

2.1 Slurring as Expressive Act

In the previous section we described some problems for effects-based accounts. In this section we describe our positive proposal: focusing on the way that slurring behavior involves expressing a derogatory attitude provides a promising path forward. According to our account, an agent *A* expresses a derogatory attitude toward a target group (or individual) *G* iff *A* performs an action such that:

- 1) *A* thereby manifests a derogatory attitude φ directed at *G* (e.g., contempt for *G*).
- 2) *A*'s action constitutes an endorsement of φ .

According to this definition, the basis for expressing a derogatory attitude for an ostensible target may in some cases be an individual trait or characteristic as opposed to their perceived group membership. Slurring acts are therefore a proper subset of acts that express derogatory attitudes. When *A* uses a slur *s* in reference to an ostensible target on the basis of their perceived group membership, *A* manifests a negative attitude toward both the ostensible target and the group of which the speaker takes the ostensible target to be a member.¹

Several theorists, including Copp (2009), Jeshion (2013), Bolinger (2017), and Camp (2018), hold that utterances of slurs signal information about a speaker's pejorative attitudes. However, because our goal here is to develop an ethical account of slurring, it is important that we be able to distinguish utterances that merely signal that the speaker has a derogatory attitude from utterances whereby the speaker *action-expresses* a derogatory attitude. Highlighting the role of endorsement enables us to explain the asymmetry between slurring and acts that merely provide descriptive information about one's inner mental life.

To see this, think about how an important first step toward changing one's objectionable attitudes is often admitting that one has them. An implicitly homophobic speaker who at the conclusion of a therapy session utters 'I hate gay people' as a neutral and well-informed evidential self-report on her mental life does not thereby action-express hatred (Bar-On 2015: 199). This speaker may consistently disavow her own objectionable attitudes and make a commitment to change them. In general, acts that merely provide evidence of an agent's harboring an attitude φ allow the agent to distance herself from the attitudes signaled (Saka 2007: 148).

On our view, the difference between self-reports (which involve mere signaling) and expression lies in the fact that the latter involves agents both manifesting and endorsing the attitudes they express. Consider a bigoted speaker whose utterance of 'I hate gay people' functions to action-express his psychological state. Someone

¹ Speakers may also use slurs metaphorically, to compare an ostensible target who is believed not to be a member of the conventional target group with members of that group (e.g., calling a man a gendered slur for women). Here, again, the speaker manifests and endorses a derogatory attitude toward both the ostensible target and the target group. For more on metaphorical uses of slurs, see Jeshion (2013: 253).

who expressively endorses such a commitment to hatred cannot consistently disavow it in the same breath. This is why the distinction between semantic expression and action-expression is so important. Though the patient who self-reports and the bigoted speaker who slurs have uttered the same sentence and have thereby semantically expressed the same proposition, the actions they are thereby performing are very different. Our moral evaluations of them should be different as well.

What about a nonracist who uses slurs in an effort to fit in with his racist friends though he does not possess the relevant attitudes? Intuitively, his slurring acts still seem expressive of derogatory attitudes, and morally objectionable, even though it is not his own attitudes to which he is giving voice. A satisfactory ethical account should be able to explain why it seems *pro tanto* wrong in general for speakers to expressively endorse an objectionable attitude whether or not they possess the attitude. It is worth taking a moment to consider how such a desideratum may be accounted for on the kind of framework we are proposing.

Some (like Sellars) explicitly restrict EXP_I to cases in which the speaker action-expresses attitudes they themselves possess (Sellars 1969: 520–21). Hence speakers putting forward attitudes they do not possess may only be described as purporting to action-express. This distinction is important in some debates (e.g., about the epistemic status of avowals), but for our purposes here it will not make a difference. The speaker who ‘purports’ to action-express an attitude φ thereby undertakes just as much commitment to, and manifests just as much endorsement of φ , as the speaker who is sincere. Compare the liar and the sincere speaker: both count as asserters who have undertaken a commitment to the truth of their utterance, and both manifest endorsement of the corresponding belief. Hence both, in the salient sense, count as having action-expressed that belief, but only the sincere speaker actually possesses it. Therefore, when we speak of ‘expression’ we refer to any speaker who, by intentionally doing something, action-expresses an attitude. This is in the spirit of Mitchell Green’s more inclusive notion of ‘expressiveness as showing a feeling . . . that need not be being felt’ (Green 2007: 194; see also Bar-On 2004: 325). One can manifest a derogatory attitude in an endorsing way, and therefore count as having expressively derogated even if one does not actually possess the derogatory attitude in question.

2.2 Defective Attitudes, Objectionable Expression

With this account of expression in hand, the next step is to establish the link between expressive act and moral wrongdoing. Outside the slurring literature, there is already a precedent for making such a connection. As Daniel Jacobson (2012) observes, in many cases the expressive aspect of an action can provide us with moral reasons not to perform it. One may be hesitant to sell one’s soul for \$2, eat a family pet, or clean a toilet with the flag of one’s country—not because one irrationally believes that doing so will have negative consequences, but because one recoils from the attitudes these actions symbolize. Jacobson concludes, ‘when we repudiate the attitude that an action expresses, then we have a perfectly good reason not to engage in it’ (2012: 313). Consider also the following variation of

an example from Kelman (1981: 34–35). An old man in Nazi Germany, who is hostile to the regime, is considering whether to publicly express support for Hitler. If he speaks out, displays the Nazi flag, performs the Nazi salute, and so on, his life and his pension will be secure. If he avoids public expressions of endorsement, official suspicion about his loyalty could put him at risk. The man is regarded as an eccentric by those who know him personally, and nobody has ever consulted his views on political questions, so what he does will have no influence on those around him. Despite the fact that publicly endorsing Hitler would have many benefits and cause no harm in this case, the man still has a very good reason not to do so.

With this precedent in mind, here is how we articulate the connection between, in particular, actions that express derogatory attitudes and moral wrongness:

Inheritance: if a derogatory attitude φ is morally defective, it is *pro tanto* wrong to action-express φ .

Inheritance offers a straightforward explanation for why slurring is wrong in solitary and solidarity cases like (A) and (B): the speakers' actions were wrong simply because the attitudes they action-expressed were morally defective.

Despite the intuitive plausibility of this principle, we still owe some justification for the claim that there are derogatory attitudes (e.g., racist beliefs, homophobic feelings, sexist perspectives, etc.) that are morally defective independent of their effects. Note that others have defended even stronger claims than what we need to endorse here. For example, Rima Basu (2019) argues that racist beliefs can constitute a wrong against the target group of the belief. A racist hermit who lives in isolation and thus will never interact with those he hates or contribute to structures that maintain institutional racism, may nevertheless wrong the target group in virtue of harboring racist beliefs about them (Basu 2019: 2504). While we are sympathetic to Basu's argument, here we need only defend a weaker claim: some derogatory attitudes are effect-independently morally defective. This leaves open the question of whether harboring such an attitude always involves wronging someone or even whether it always warrants blame.

Though the inheritance principle relies on a weaker assumption, it argues for a broader conclusion. If Basu's hermit engages in actions that express his defective attitudes, he can be blameworthy for performing such actions even if he is not blameworthy for having the attitudes. This is because expressive acts are under one's immediate control in a way that attitudes are not. The inheritance principle explains the wrongness of any act that involves the expression of a morally defective derogatory attitude, whether or not forming the original attitude was under the agent's control. Moreover, this principle allows us to explain why such acts are still wrong even if the speakers themselves do not possess the derogatory attitudes. Moral reasons to abjure an attitude are also moral reasons to refrain from expressing it.

2.3 Explaining the Moral Defectiveness of Derogatory Attitudes

On our account, the wrongness of slurring ultimately depends on the moral properties of the derogatory attitudes that slurring acts express. It is therefore necessary at least to gesture at a broader theory of what makes these attitudes morally defective. We believe that here Macalester Bell's account of inapt contempt provides a promising path forward. According to Bell, 'on a minimally acceptable morality, our moral relationships are characterized by respect and affective openness' (Bell 2013: 213). The affectively open person is one who 'will not respond to others with hostility or ill will unless she has been given good reason to do so' and is 'open to appreciating the good qualities of persons' (Bell 2013: 213–14). For Bell, contempt that is based on biases and stereotypes, like racism, is objectionable because it is inconsistent with affective openness. To be affectively closed, so to speak, is to be unresponsive to another's normatively significant characteristics. Even if the affectively closed person never has an opportunity to actually interact with members of the group against which they are biased, it can still be true that *were* the person to interact with or learn more about members of this group he *would* deny them the power to shape his attitudes.

Expanding on Bell's account, we claim that objectionable derogatory attitudes in general are morally defective precisely because agents who harbor them are thereby affectively closed—unable to see the derogated person as 'fully in possession of normative powers' (Bell 2013: 214). To be clear, a derogatory attitude may be morally objectionable even if it involves a non-moral normative judgment. For example, consider perspectives that hold fat people to be unintelligent, regard gay people as rude, or consider women to be physically weak. These are quite clearly morally defective derogatory attitudes even though they involve ranking groups according to non-moral norms (rationality, etiquette, and athleticism, respectively). The reason these attitudes are morally defective is that harboring them makes one affectively closed. Bell explains the lack of openness by appealing to a variety of 'aptness conditions' (Bell 2013: 147–51). While several of these conditions apply only to Bell's discussion of contempt, we think two broad categories of criteria are useful for understanding the affective closure that comes from morally objectionable derogatory attitudes in general.

On the one hand, a derogatory attitude makes one affectively closed and is thereby morally defective if it is morally indefensible. As Bell points out, 'moral theorists will differ with regard to what they think counts as an indefensible reason for contempt' (Bell 2013: 148). With Bell, we assume that a wide range of deontological and virtue ethics are capable of accounting for the moral defectiveness of affective closure. While the moral framework that Bell herself develops gives central place to virtues and vices, it is also heavily inspired by the work of T. M. Scanlon (Bell 2013: 83–84, 213). However, even from a theory neutral perspective, we can at least identify some clear cases. Derogatory attitudes based on biases and stereotypes are paradigm cases of attitudes that cannot be morally justified even if they seem epistemically justifiable (e.g., stereotypes that are statistically likely to be true due to correlations). Derogatory attitudes that are unresponsive to moral considerations in favor of revision (e.g., forgiveness, reconciliation, extenuating circumstances, etc.) are also morally indefensible (Bell 2013: 149–50).

On the other hand, harboring a derogatory attitude may make one affectively closed (and hence morally deficient) in virtue of its being epistemically indefensible. If the target/group does not really manifest the fault in question, if the derogator's attitude is based on reasons for which he lacks evidence, or if the derogator is unwilling to revise his attitudes in the face of new evidence, then these are good grounds for diagnosing affective closure (Bell 2013: 147–48). Agents who exhibit derogatory attitudes in such cases thereby fail to grant targets 'the respect and affective openness characteristic of unimpaired moral relations' (Bell 2013: 213). Derogatory attitudes that base normative evaluations on irrelevant characteristics or prejudices are likely to be both morally and epistemically indefensible (Appiah 1990; Basu and Schroeder 2019).

Let us sum up the account so far: slurring speakers action-express morally objectionable derogatory attitudes, thereby endorsing affectively closed perspectives. Our inheritance principle extends Bell's account to explain the effect-independent moral wrongness of slurring acts (including the solidarity and solitary examples discussed above). Because it is *pro tanto* wrong to endorse or commit oneself to an affectively closed perspective, whenever slurring involves such an act of expressive commitment it is *pro tanto* wrong.

Let us now return to an issue raised in section 2.1: consider a slurring act that is performed by an agent who does not themselves possess the relevant affectively closed derogatory attitudes. How does our theory explain the wrongness of slurring in these cases? Another example will be instructive. Imagine Jack pledges allegiance to a white supremacist organization. Jack understands that the organization is committed to this racist ideology but does not personally approve of its views. Thus, Jack is not himself affectively closed with respect to people the organization regards as inferior. Jack's motivation to join is just that he hopes to make advantageous professional connections. All else being equal, it is *pro tanto* wrong for Jack to engage in acts of endorsement of this sort and thereby commit to being affectively closed even if he does not intend to follow through.

We qualify the wrong Jack commits as *pro tanto*, because speakers who derogate may sometimes be all-things-considered justified in doing so. Imagine that Jack was actually an undercover agent who infiltrated the white supremacist organization as part of an effort to undermine and dismantle it. Despite not sharing the racist attitudes of the other members, Jack must engage in slurring behavior from time to time in order to avoid detection. On our view, these slurring acts constitute endorsements in the sense that they are expressive of racist commitments (albeit commitments Jack does not actually possess). The reason why these acts are effective in preserving Jack's cover and ingratiating him with the other members of the group is precisely that they are expressive acts of commitment. While the wrongness of Jack's slurring behavior persists, his actions are all-things-considered justified by the value of working toward destroying the organization.

Thus, our view predicts that the moral objectionableness of slurring can be outweighed but not defeated—it is *pro tanto* wrong rather than *prima facie* wrong. We believe this is the right result, as it would be morally intelligible and appropriate for Jack subsequently to regret having had to say and do the racist

things he did as part of his mission even though he is not all-things-considered blameworthy (and is probably even praiseworthy) for what he has done.

3. Virtues of the Expression-Centered Account

Some readers may feel that recognizing the effect-independent wrongness of slurring acts makes little practical difference. After all, in the real world, does slurring not almost always risk causing negative effects of some sort? While it is true that slurring often does so, we believe it is dangerous to place all our ameliorative eggs in the measurable-effects basket. Doing so opens the door to slurring agents who defend their actions by claiming that they are not prejudiced themselves, that they ‘are not harming anyone,’ or that listeners ‘should not be so sensitive’. Our expression-centered account has the resources to shift the burden of proof from the individuals and groups targeted by derogatory language and condemn derogatory expressive acts on the grounds that affective openness is owed to all.

Inheritance’s ameliorative benefits are thus in the spirit of Kate Manne’s recent call to ‘make misogyny more epistemically tractable’ (Manne 2018: 60). Not only is it a moral failing to *be* affectively closed, it is also morally wrong to expressively endorse attitudes that *would make one* affectively closed. Instead of needing to read a speaker’s mind to see whether he possesses the relevant derogatory attitudes or needing to prove that someone has suffered, all we must know in order to condemn a slurring act morally is whether a member of the target group could reasonably perceive the action as expressive of a morally objectionable derogatory attitude. Protestations like ‘I am not a sexist’ or ‘I am not a racist’ are no defense: derogatory behavior that is expressive of objectionable attitudes still functions to oppress and exclude by projecting an affectively closed perspective. And such behavior remains morally defective even when enacted by speakers who do not have bad intentions. Even if we hesitate to always describe the accidental derogator as vicious or blameworthy, our account has the resources to explain the moral reasons against what he has done and in favor of changing his behavior.

As noted above, even in cases where effects are produced, they may sometimes be very minimal. For example, in communities where bigots represent a small minority and are well known, their slurring utterances may be easily anticipated and quickly repudiated by the majority of hearers. The harm that bigots are able to produce in such communities is relatively small. Nevertheless, their behavior is still wrong, and this wrongness cannot be fully captured by reference to measurable negative effects. What justifies the moral criticism of these bigots need not be some horrible effect they have successfully produced (there might be none), but can instead be the moral defectiveness of the bigoted attitudes they thereby express. In fact, it is precisely this strong, unilateral, justified denunciation that often serves to sap slurring utterances of their ability to harm. According to an effect-exclusive or harm-based account, the act of morally criticizing a bigot’s use of a slur in such cases would have the paradoxical status of removing its own justification. This leads to a kind of reverse self-effacing problem: the theory recommends acting as if the bigot’s speech is likely to cause harm when actually it is not. This highlights

another strength of our approach: it can offer a full-throated justification of the practice of condemning morally objectionable expressive acts.

Though our focus in this paper has been on slurring acts, an important puzzle remains: the moral status of merely mentioning slur words. One commonly observed feature of slur words is their wide-scoping behavior under embedding. Utterances of slurs can be offensive and objectionable despite being embedded within quotation marks or within the scope of truth-conditional operators like negation (Saka 2007: 122). This is because slur words can retain their status as conventional expressive vehicles—vehicles that enable speakers to enact endorsement—even when embedded.

A good example of this can be seen in an event at a Donald Trump rally in New Hampshire in 2016. While Trump was criticizing fellow Republican presidential primary candidate Ted Cruz, a supporter near the front of the crowd shouted a slur. In response, here is what Trump said:

She just said a terrible thing. You know what she said? Shout it out because I don't want to. Ok. You're not allowed to say, and I never expect to hear that from you again. She said—I never expect to hear that from you again! She said he's a 'pussy!' Terrible. Terrible. That's terrible. (Spoken Word 2016).

Here Trump embeds the slur in a quote, but his utterance is offensive and objectionable nevertheless. Clearly, embedding does not necessarily isolate a speaker from the ethical implications of uttering slurs. Thus, the inheritance principle applies in metalinguistic contexts as well as in standard uses. The act of mentioning a slur can, in the right context, constitute a slurring act and therefore an act of endorsing an objectionable derogatory attitude.

Of course, uttering a slur word need not always constitute endorsement. We expect that readers of this article will not interpret us as endorsing Trump's attitudes simply by quoting him above. Likewise, it would be unreasonable to accuse a speaker of slurring merely because she uttered a slur while directly quoting someone as part of her testimony in court (Bolinger 2017: 442). Beyond such cases of direct quotation, there are various pedagogical contexts wherein a speaker's mentioning utterance of a slur seems permissible. These include cases where one intends to teach someone who is unfamiliar with the word, say a non-native speaker or child, that it ought not be used. The permissibility of such utterances will depend, *inter alia*, on the speaker's being sincere about discouraging the use of the word. Part of the reason that Trump's utterance is so morally problematic is that the context makes clear that his mentioning of the slur serves no worthwhile academic or pedagogical purpose. His own history of using offensive language, expressing derogatory attitudes toward women and other oppressed groups, and denigrating his political opponents belies his protestations.

Utterances of slurs in academic contexts—colloquia, journal articles like this one, seminar discussions, and so on—are fairly common as a result of the large body of literature on slurs that has emerged over the last decade. Those who attend academic talks about slurs may get the impression that some speakers feel they ought to avoid

uttering slurs because even mere mentions can be controversial. However, other speakers sometimes feel the need to mention a slur in full (rather than employ a circumlocution). In one recent incident, a creative writing professor at The New School, Laurie Sheck, asked students whether they had seen the 2016 documentary about James Baldwin, *I Am Not Your Negro*. Sheck explained that the documentary's title comes from a statement made by Baldwin, in which Baldwin uttered the n-word (instead of 'negro'). As part of her explanation, Sheck mentioned the n-word in its uncensored form (Flaherty 2019). In response to a white student's complaint, the university investigated Sheck for potentially violating its antidiscrimination policy (though they quickly cleared Sheck of wrongdoing).

We believe that our account has the resources to provide at least some action guidance with respect to this issue. There are several theoretical goals that require quoting slurs in full. However, the value of accomplishing these goals must always be weighed against the disvalue of mentioning the slur. If what we have said above about the ethics of slurring is true, then there are two sources of disvalue that must be taken into account: the wrong of enacting endorsement of an affectively closed perspective and the wrong of causing problematic effects. In many cases, despite the value of goals that could be achieved by doing so, mentioning the slur will not be all-things-considered justified.

For example, mentioning slurs may be crucial to taxonomizing them. If one claims, as a number of authors have, that some pejoratives are metaphors (Allen 1983; Goff et al. 2008; DiFranco 2015; Camp 2017), it is reasonable to expect one to provide an unabbreviated and uncensored example—sometimes a circumlocution will simply not do. However, for this purpose one example should suffice, and an example of a less objectionable slur at that. Because accomplishing these goals does not require mentioning particularly vicious slurs like the n-word, quoting such words may rarely (perhaps never) be justified by academic or pedagogical goals. Insofar as it is unnecessary to utter the most objectionable derogatory words, hearers may reasonably perceive speakers who *do* mention them as action-expressing morally defective derogatory attitudes.

Also consider Camp's (2013) observation that including uncensored slurs in examples may sometimes be necessary for evoking in the audience the kind of experiential states one is attempting to theorize about. Camp includes a disclaimer:

I am going to mention (though not use) a variety of slurs in contemporary use. This will offend some readers . . . but I believe we can understand slurs' actual force only by considering examples where we ourselves experience their viscerally palpable effects. I hope the offense is offset by a commensurate gain in understanding. (Camp 2013: 331).

This certainly seems plausible. Consider metaphorical expressions whose derogatory power seemingly depends on evoking imagery. It is easy to see how interpreting 'Naomi mopped the floor with George during their debate' may mandate imagery formation for hearers unfamiliar with this metaphor. Forming a mental image of

Naomi holding George upside down and pushing him across the debate stage floor so as to mop it with his hair may help a hearer who is initially confused grasp the metaphor for intellectual domination.

Likewise, we can imagine audience members who are not competent with a metaphorical slur needing to form a derogatory image to fully appreciate what is wrong with using it. Being momentarily sullied by hearing and interpreting a metaphorical slur and (as a result) forming a mental image, may help such listeners appreciate what makes the slur so rhetorically powerful. However, *pace* Camp (2017), we do not think that interpreting metaphorical insults requires actually adopting a derogatory perspective on the target. We are hesitant to ascribe such overwhelming powers to metaphor. While many metaphors do necessitate imagery formation, it should be possible to form such a pejorative image without thereby endorsing it.

Though we agree with Camp that gains in understanding sometimes require experiencing visceral effects, it is also important not to homogenize the audience (or the theorists) implicitly. In order for the pedagogical value of mentioning a slur to justify uttering it, we believe that three facts must be firmly in common ground: (1) the interlocutors are participating in an academic exercise, (2) everyone present understands the nature and value of this exercise, and (3) the audience is committed to quarantining the objectionable imagery they are forming when interpreting the slur. However, if the audience includes members of the slur's target group who are already familiar with the word's rhetorical effects, then chances are they would not benefit from hearing the slur. In such a case, extra caution is warranted: well-intentioned allies may unintentionally slur when their repeated utterances of these words could be reasonably interpreted by target group members in the audience as expressive of a lack of respect or consideration. Hence, the duty to project affective openness counts against mentioning the slur.

At the same time, in other contexts where it seems clear that the speaker will be able to avoid endorsing a derogatory attitude, the possibility of causing negative effects can still count against mentioning the slur—utterances of slurs can still produce effects like triggering long-standing trauma. While speakers who give examples of slurs in academic contexts are not themselves slurring (i.e., they are not thereby expressing approval of the derogatory imagery these slurs characteristically evoke), this fact alone does not magically license all utterances. The pedagogical value of uttering a slur depends on who the audience is. If the audience is made up of both incompetent speakers and members of the slur's target group, it is important not to assume that the goal of understanding is the only one that matters. In such cases, the testimony of members of the target group should be sufficient to establish a slur's status for the purpose of discussion.

In light of this, what norms should academics who discuss slurs adopt in order to project affective openness? According to our account, if one is confident that an abbreviated or censored version of a slur will be sufficient to bring to mind the example one wants to discuss, then one ought to mention only that version in examples (rather than the uncensored version). If not, then a single uncensored occurrence should be sufficient for drawing the audience's attention to the case one wants to discuss and a circumlocution ought to be used thereafter. It should

also always be made clear what the author's specific pedagogical aim is in mentioning a slur in order to avoid manifesting an affectively closed perspective (for a related point about how offense is generated, see Bolinger 2017: 452).

4. Concluding Remarks on the Potential for Morally Valuable Slurring

Some theorists have understood slurring as a form of dehumanization (Jeshion 2013: 232; Smith 2011; Gaita 1998). If they are right, then harboring derogatory attitudes toward groups and expressing them is always wrong insofar as dehumanization is always wrong. However, we believe that the dehumanization model defines slurring acts and derogatory attitudes too narrowly. Certainly, thinking of a racial group as cockroaches is a form of dehumanization. However, Kate Manne (2018) has argued that some morally objectionable attitudes, attitudes that form the basis for mistreatment or subjugation, can presuppose the target's humanity. A man who demands a woman's love and affection—and responds violently in retaliation when she refuses to provide it—implicitly acknowledges her agency and autonomy: particularly her subjectivity, preferences, and capacity to form deep emotional attachments (Manne 2018: ch. 5).

Feeling contempt for an individual on the grounds that he is racist or sexist is also compatible with recognizing his humanity; morally justified contempt of this sort need not dehumanize its target. Moreover, thinking of those who deliberately violate moral norms as having a lower moral standing may motivate communities to seek justice and put perpetrators in a position to appreciate the reasons they have to change their ways (Bell 2013: 225). As long as the perspective is grounded in the morally relevant features and behaviors of the target, it is consistent with our obligation to be affectively open. At the same time, if Bell (2013) is correct in claiming that features like race and gender are irrelevant to a group's normative status, then expressing derogatory attitudes toward individuals on the basis of their membership in these groups will always involve committing oneself to an affectively closed perspective (and thus will never be fitting or just). It would seem then that slurring in particular can never be morally valuable.

Can we find exceptions to Bell's rule? Consider a member of a marginalized group who harbors negatively valenced attitudes toward a powerful majority group that occupies an unjust social position of power. According to some views, categories such as race and gender are socially constructed in such a way that they are not morally neutral (Haslanger 2000). On such accounts, for example, occupying a dominant position in a social hierarchy is part of what it means to be white. Those who take such hierarchies to be unjust may consequently cultivate derogatory attitudes toward whiteness itself. In the context of a normative understanding of these categories, such attitudes may be morally justified. 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.

However, here it is important to distinguish between attitudes directed toward the category, *whiteness*, and attitudes directed toward individual members. In some cases, satisfying one's obligation to be affectively open toward individual members

of a category may be compatible with harboring a fitting and morally valuable derogatory attitude toward the category itself. Of course, the value of using slurs to express justified condemnations of dominant groups and foster solidarity among marginalized communities must still be weighed against the potential harms to individuals that using these slurs could cause.

For those interested in understanding the ethical implications of slurring, recognizing its expressive quality is essential. Effects-based approaches can count many intuitively objectionable cases as wrong, but not all. In this article, we proposed an expression-centered account of the wrongness of slurring that is explanatorily more fundamental. Solitary and solidarity slurring that causes no harm or psychological effects can still be judged *pro tanto* wrong because the speaker's expressive act inherits the moral defects of the attitudes they express. In turn, agents who possess affectively closed derogatory attitudes count as morally criticizable, regardless of whether they cause harm, for failing to accord others the affective openness they are owed. Our view can also explain why slurring has the potential to bring about its characteristic negative effects even when speakers do not possess the attitudes they express: problematic effects result from enacting endorsement of affectively closed perspectives.

In addition to accounting for our intuitive moral assessment of causally inert slurring, the expression-centered account can also shed light on morally valuable derogatory expressive acts. According to the principle of inheritance, derogatory expressive acts are only as objectionable as the attitudes thereby manifested and endorsed. The fact that a particular derogatory attitude toward a vicious person, oppressive group, or unjust institution is fitting and morally valuable can explain why expressions of that attitude may sometimes be permissible as well.

RALPH DIFRANCO
AUBURN UNIVERSITY
rcd0034@auburn.edu

ANDREW MORGAN
UNIVERSITY OF ALABAMA IN HUNTSVILLE
adm4yd@virginia.edu

References

- Allen, I. (1983) *The Language of Ethnic Conflict: Social Organization and Lexical Culture*. New York: Columbia University Press.
- Anderson, L. (2015) 'Racist Humor'. *Philosophy Compass*, 10, 501–509.
- Anderson, L., and E. Lepore. (2013) 'What Did you Call Me? Slurs as Prohibited Words'. *Analytic Philosophy*, 54, 350–63.
- Appiah, K. A. (1990) 'Racisms'. in D. Goldberg (ed.), *Anatomy of Racism* (Minneapolis, MN: University of Minnesota Press), 3–17.
- Ashwell, L. (2016) 'Gendered Slurs'. *Social Theory and Practice*, 42, 228–39.
- Bar-On, D. (2004) *Speaking My Mind: Expression and Self-Knowledge*. Oxford: Oxford University Press.

- Bar-On, D. (2015) 'Expression: Acts, Products, and Meaning'. In S. Gross, N. Tebben, and M. Williams (eds.), *Meaning Without Representation: Expression, Truth, Normativity, and Naturalism* (Oxford: Oxford University Press), 180–209.
- Basu, R. (2019) 'The Wrongs of Racist Beliefs'. *Philosophical Studies*, 176, 2497–2515.
- Basu, R., and M. Schroeder. (2019) 'Doxastic Wronging'. In B. Kim and M. McGrath (eds.), *Pragmatic Encroachment in Epistemology* (New York: Routledge), 181–205.
- Bell, M. (2013) *Hard Feelings: The Moral Psychology of Contempt*. Oxford: Oxford University Press.
- Bolinger, R. J. (2017) 'The Pragmatics of Slurs'. *Nous*, 51, 439–62.
- Camp, E. (2013) 'Slurring Perspectives'. *Analytic Philosophy*, 54, 330–49.
- Camp, E. (2017) 'Why Metaphors Make Good Insults: Perspectives, Presupposition, and Pragmatics'. *Philosophical Studies*, 174, 47–64.
- Camp, E. (2018) 'A Dual Act Analysis of Slurs'. In D. Sosa (ed.), *Bad Words: Philosophical Perspectives on Slurs* (Oxford: Oxford University Press), 29–59.
- Copp, D. (2009) 'Realist-Expressivism and Conventional Implicature'. *Oxford Studies in Metaethics*, 4, 167–202.
- DiFranco, R. (2015) 'Do Racists Speak Truly? On the Truth-Conditional Content of Slurs'. *Thought: A Journal of Philosophy*, 4, 28–37.
- DiFranco, R. (2017a) 'Derogation Without Words: On the Power of Non-Verbal Pejoratives'. *Philosophical Psychology*, 30, 784–808.
- DiFranco, R. (2017b) 'Appropriate Slurs'. *Acta Analytica*, 32, 371–84.
- Feinberg, J. (1985) *Offense to Others: The Moral Limits of the Criminal Law*. Oxford: Oxford University Press.
- Flaherty, C. (2019) 'New School Drops N-Word Case'. *Inside Higher Ed*. August 19, 2019. <https://www.insidehighered.com/news/2019/08/19/creative-writing-professor-cleared-discrimination-saying-slur-quoting-james-baldwin>.
- Gaita, R. (1998) 'Racism: The Denial of a Common Humanity'. In R. Gaita (ed.), *A Common Humanity: Thinking About Love and Truth and Justice* (New York: Routledge), 57–72.
- Gimbel, S. (2018). *Isn't That Clever: A Philosophical Account of Humor and Comedy*. New York: Routledge.
- Goff, P. A., J. L. Eberhardt, M. J. Williams, and M. C. Jackson. (2008) 'Not Yet Human: Implicit Knowledge, Historical Dehumanization, and Contemporary Consequences'. *Journal of Personality and Social Psychology*, 94, 292–306.
- Green, M. (2007) *Self-Expression*. Oxford: Oxford University Press.
- Haslanger, S. (2000) 'Gender and Race: (What) Are They? (What) Do We Want Them To Be?'. *Nous*, 34, 31–55.
- Hay, R. (2013) 'Hybrid Expressivism and the Analogy between Pejoratives and Moral Language'. *European Journal of Philosophy*, 21, 450–74.
- Hom, C. (2012) 'A Puzzle About Pejoratives'. *Philosophical Studies*, 159, 383–405.
- Hooker, B. (2015) 'Rule Consequentialism'. In E.N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/consequentialism-rule/>.
- Jacobson, D. (2012) 'Moral Dumbfounding and Moral Stupefaction'. *Oxford Studies in Normative Ethics*, 2, 289–315.
- Jay, T. (2009) 'The Utility and Ubiquity of Taboo Words'. *Perspectives on Psychological Science*, 4, 153–61.
- Jeshion, R. (2013) 'Expressivism and the Offensiveness of Slurs'. *Philosophical Perspectives*, 27, 231–59.
- Jones, T., and E. E. Wade (2020) 'Me Too? Race, Gender, and Ending Workplace Sexual Harassment'. *Duke Journal of Gender Law & Policy*, 27, 203–25.
- Kandil, C. Y. (2020) 'Asian Americans report over 650 racist acts over last week, new data says'. *NBC News*. March 26, 2020. <https://www.nbcnews.com/news/asian-america/asian-americans-report-nearly-500-racist-acts-over-last-week-n1169821>.
- Kelman, S. (1981) 'Cost-Benefit Analysis: An Ethical Critique'. *Regulation*, 5, 33–40.
- Mag Uidhir, C. (2013) 'What's so Bad about Blackface?'. In D. Flory and M. Bloodsworth-Lugo (Eds.), *Race, Film, & Philosophy* (New York: Routledge), 51–68.

- Manne, K. (2018) *Down Girl: The Logic of Misogyny*. Oxford: Oxford University Press.
- Morgan, A. (2019) 'Plato's Revenge: Moral Deliberation as Dialogical Activity'. *Pacific Philosophical Quarterly*, 100, 69–89.
- Richter, R. (1986) 'On Philips and Racism'. *Canadian Journal of Philosophy*, 16, 785–94.
- Saka, P. (2007) *How to Think About Meaning*. Berlin: Springer.
- Sellars, W. (1969) 'Language as Thought and as Communication'. *Philosophy and Phenomenological Research*, 29, 506–27.
- Sennet, A., and D. Copp. (2015) 'What Kind of Mistake is it to Use a Slur?'. *Philosophical Studies*, 172, 1079–1104.
- Smith, D. L. (2011) *Less Than Human: Why We Demean, Enslave, and Exterminate Others*. New York: St. Martin's Press.
- Spoken Word. (2016) 'Ted Cruz is a Pussy—Trump Rally Goer Calls Cruz Pussy'. *YouTube*, 8 February 2016. <https://www.youtube.com/watch?v=XX4MO7YHqeg>.
- Tirrell, L. (2017) 'Toxic Speech: Toward an Epidemiology of Discursive Harm'. *Philosophical Topics*, 45, 139–61.