When Code Words Aren't Coded

Patrick O'Donnell
Oakton Community College
podonnell@oakton.edu

Forthcoming 2020, *Social Theory and Practice
Please refer to the final version for citation purposes

Abstract: According to the "standard framing" of racial appeals in political speech, politicians generally rely on coded language to communicate racial messages. Yet recent years have demonstrated that politicians often express quite explicit forms of racism in mainstream political discourse. The standard framing can explain neither *why* these appeals work politically nor *how* they work semantically. This paper moves beyond the standard framing, focusing on the politics and semantics of one type of explicit appeal, candid racial communication (CRC). The linguistic vehicles of CRC are neither true code words, nor slurs, but a conventionally defined class of "racialized terms."

Keywords: Code words, racialized terms, candid racial communication, populism, political speech, norm of racial equality

> "[Race talk is] the explicit insertion into everyday life of racial signs and symbols that have no meaning other than pressing African-Americans to the lowest level of the racial hierarchy... There is virtually no movement up-- for blacks or whites, established classes or arrivistes-- that is not accompanied by race talk. Refusing, negotiating, or fulfilling this demand is the real stuff, the organizing principle of becoming an American."
>
> --Toni Morrison, "On the Backs of Blacks," *Time*, December 2, 1993

> "Let them call you racists. Let them call you xenophobes. Let them call you nativists. Wear it as a badge of honor."
>
> --Steve Bannon, Speech to the French National Front, March 11, 2018

There is political utility in the racist thoughts and feelings of a citizenry. Racism can motivate us to elect certain candidates, support certain policies, and even go to war. Yet Americans, along with most citizens of liberal democratic states, tend to disapprove of racism. Openly espousing racist views and policies is thus out of the question for politicians who want to win friends and influence people. In order to tap into the political potential of racism, they must turn to other means.

One of those other means is language which exploits racist sentiments without appearing racist itself. In a 1981 interview, renowned Republican strategist Lee Atwater described this tactic in

the context of the of the Southern Strategy, in which the Republican party successfully consolidated a loyal base of Southern white voters partially by appealing to their racial resentments:

> You start out in 1954 by [slurring blacks.] By 1968 you can't [slur blacks]—that hurts you, backfires. So you say stuff like, uh, forced busing, states' rights, and all that stuff, and you're getting so abstract. Now, you're talking about cutting taxes, and all these things you're talking about are totally economic things and a byproduct of them is, blacks get hurt worse than whites...if it is getting that abstract and that coded, we're doing away with the racial problem one way or another... "We want to cut this," is much more abstract than even the busing thing, and a hell of a lot more abstract than [slurring blacks.] (recorded interview, quoted in Perlstein 2012.)

At the center of this sleight-of-hand were *racial code words*. Terms like "forced busing," "state's rights," and even "cutting taxes" carry a surface message devoid of any obvious racial reference while also carrying a "coded" racial message. The function of these code words is to exploit an audience's racial fears and antipathies while ostensibly complying with a general "norm of racial equality" shared by politicians and constituents alike.

Extant work on the role of code words in political communication assumes that most public racial appeals in the post-Civil Rights era have roughly the same structure. According to this "standard framing," i) the norm of racial equality constrains contemporary political discourse. The hegemony of this norm explains why ii) racial appeals in politics are generally *implicit* rather than *explicit* and why iii) *code words*, as opposed to slurs, are the linguistic vehicles of racial appeals.

Yet the last few years have made it clear that a good deal of public racial discourse does not fit well within the standard framing. In the wake of a global populist mobilization-- including the "Brexit" referendum, the rise of Donald Trump to the American presidency, the high-profile French presidential campaign of Marine Le Pen, and the successful campaign for the Austrian chancellorship led by Sebastian Kurz--the public acceptability of discourse which *explicitly* targets racialized groups for exclusion and marginalization is on the rise.[1] Speakers do not rely on "coded" language in getting their race-specific messages across, and audiences don't find themselves so conflicted by agreeing to the content of race-specific messages.

I think that this shift in the tenor of public racial discourse should be taken at face value. The standard framing can no longer be regarded as the only game in town in understanding the mechanics of mainstream racial communication. The negative goal of this paper is to show that the standard framing gives unsatisfying accounts of the *normative structure* of these openly race-specific appeals as well as the *meaning* of their linguistic vehicles. We will also see that the standard framing faces substantial empirical and conceptual problems quite independently of this explanatory limitation.

[1] Here the "public acceptability" of this discourse should be understood in a descriptive sense, not a normative one. Dispositions to accept explicitly racialized discourse are on the rise, whether or not such discourse is acceptable in a normative sense.

The positive goal is to begin the work of theorizing racialized political communication beyond the ambit of the standard framing. We will explore some examples of what I call *candid racial communication*. Candid racial communication (CRC) is interesting precisely because it thrives in communities where norms of racial equality are not actively observed. Many of the core linguistic vehicles of CRC are interestingly different from the terms usually described as "code words." Such terms belong to a "not-so-coded" category of expressions I will call *racialized terms*.

Section 1 presents the standard framing in more detail. Section 2 shows that the standard framing has conceptual and empirical limitations that hinder its explanatory generality. Section 3 presents some examples of candid racial communication, and shows that CRC openly flouts those norms which the standard framing assumes are generally accepted among Americans. The section also includes a discussion of the relevance of CRC to the rhetorical strategies of right-wing populism. Sections 4 and 5 show that the normative structure of CRC is relevant to determining what racialized terms mean in context. I argue that in the context of CRC, key terms such as "thug," "criminal alien," "terrorist," and "immigrant" take on specifically racialized referential contents. We should thus understand racialized terms as not quite "coded" and not quite "slurring," but as instances of "community-specific speech" in which terms are conventionally repurposed to refer to racialized groups in a way that is both explicit and acceptable by the lights of that community's racial discourse norms.

1. The Standard Framing

The standard framing has two components. The first is a communication theory which specifies the conditions under which racial appeals can be effective. The second is a semantic theory which specifies how code words in particular allow interpreters to pick up on substantive race-specific messages.

1. 1. The Communication Theory

Tali Mendelberg's (2001) groundbreaking study of the function of racial appeals in American politics is largely responsible for the core assumptions of the standard framing. Mendelberg's core insight is that the character of racial appeals varies depending on the content of generally accepted norms involving race. In the Jim Crow South, the operative racial norm "dictated conformity to the basic precepts of white superiority and black inferiority" (Mendelberg 2001: 28). Under the sway of this "norm of racial inequality," political candidates could garner support among their white constituencies through *explicit* racial appeals which reflected commitment to white supremacy. Politicians routinely used slurs, made explicit reference to anti-black stereotypes, and continually referenced blacks in the service of playing on white anxieties about the social consequences of desegregation.

Yet a new racial consensus emerged in the post-Civil Rights era. In the place of a norm of racial *inequality* comes a norm of racial *equality*: "The norm of racial equality is the consensus that the ideology of white supremacy is morally and empirically bankrupt. The norm repudiates the notion

that blacks are inalterably *[sic]* inferior and rejects this idea as a justification for treating blacks less favorably than whites" (Mendelberg 2001: 112). The hegemony of this new norm means that Americans are reluctant to support speech or sentiments that are *explicitly* racist, and they are reluctant to think of themselves as racist. Flouting the norm in the political arena can thus be politically costly, since it might alienate those even weakly committed to it.

The technology of *implicit* racial appeals solves this strategic problem. Explicit tirades about black indolence, sexual deviancy, and criminality are transformed into sermons on the dangers that "welfare" poses to the "work ethic," warnings about the corrosive effects of the degeneration of "the family," and demonization of "thugs" and "super predators" who could do with a healthy dose of "law and order." Since none of these terms makes explicit reference to race, they afford their speakers a certain degree of plausible deniability: if accused of violating the norm of racial equality, speakers can point to the fact that terms like "thug" and "welfare" do not obviously have any racial content. Similarly, such terms allow audiences--even those strongly committed to the norm-- to accept implicit racialized messages without thinking that they *themselves* are violating the norm.

To see this dynamic at work, consider a comment made by then-US Representative Paul Ryan during a radio interview:

> [We] have got this tailspin of culture in our inner cities, in particular, of men not working and just generations of men not even thinking about working or learning the value and the culture of work; and so there's a real culture problem here that has to be dealt with. (Kertscher 2014)

The anchoring terms of this discourse-- "inner cities," "culture," "(not) working"-- make no explicit reference to race. Yet each is plausibly an instance of coded speech which might be reasonably expected to leave the following message with the listener: *black men in urban centers are lazy and unwilling to work, and would rather subsist on government handouts*. Ryan was excoriated for his comments. California Representative Barbara Lee bluntly declared "Let's be clear, when Mr. Ryan says 'inner city,' when he says 'culture,' these are simply code words for what he really means: 'black.'" (Blow 2014) Yet since Ryan had used terms which could in principle apply to *any* racial demographic, he could claim in his defense that "there was nothing whatsoever about race in my comments at all-- it had nothing to do with race." (Blow 2014)

Mendelberg argues that when the norm of racial equality is dominant, implicit racial appeals lose their effectiveness when the underlying racial character of the appeal is made explicit. Supporting this plank of her theory is the book's central case study: the "Willie Horton" script. Horton was a black convict who assaulted a white couple after fleeing during a weekend furlough from a Massachusetts state prison. In the months leading up to the 1988 Presidential Election, the Bush campaign used the Horton episode to suggest that opponent Michael Dukakis, who as governor of Massachusetts had supported the furlough program which allowed Horton's escape, was "soft on crime." The most famous use of the Horton script was the "Weekend Passes" television ad, which ran on television from September 21 to October 5, 1988. Behind a voiceover

summarizing his crimes, the ad showed two images of Horton (tall, bearded, and sporting an Afro), one a scowling mugshot, and one of him being escorted from a courtroom by police. Neither the text of the ad, nor most of the subsequent media references to the Horton script, mentioned race.

Yet on October 21, the Rev. Jesse Jackson publicly claimed that the Horton script was racist, and accused the Bush campaign of "playing the race card" in a way that capitalized on white racial resentment. Drawing on her analysis of contemporaneous news coverage of the issue and data on voter preferences before and after Jackson's accusation, Mendelberg argues that the Horton script lost a good deal of its rhetorical force after its racial dimensions were made explicit. Once white Americans became aware of the fact that their racial resentments were being primed by the script, they saw it as violating the norm of racial equality and disavowed commitment to its previously implicit message.

1.2 The Semantic Theory

As the linguistic vehicles which carry implicit racial appeals, the central function of code words is to "open up space for deniable norm violations" (Khoo 2017) so that interlocutors can traffic in implicit racial messages while believing that their thought and talk is in accordance with the norm of racial equality. A semantic theory of code words must explain what it is about code words that allows them to perform this functional role.

Recently, Justin Khoo (2017) has proposed an elegant "simple theory" of code words. The core virtue of Khoo's account over extant competitors is that it does not require us to embrace any substantive assumptions about the nature of linguistic meaning. In describing how code words work, all we need to assume is that code words somehow prime stereotypes and resentful feelings about racialized groups. Consequently, the race-specific reading of a sentence such as "we have a culture problem in our inner cities" is not *communicated* by virtue of the sentence's encoded semantic meaning, but *inferred* from or *associated* with the term itself by an interpreter.

According to Khoo, the effects of code words can thus be modelled as follows:

Pre-existing belief: The inner city is populated by poor African-Americans.
Code word sentence: "We have a culture problem in our *inner cities*."
Racial inference: The problems in the inner city have to do with the cultural shortcomings of African-Americans.

More generally, where C is a code word, R is a racialized feature, and x is an individual or policy under discussion, C prompts the following sort of inferences:

EXPLICIT STATEMENT: x is C.
EXISTING BELIEF: If something is C, then it is R
INFERRED: x is R.

This is already enough to explain how code words open up deniable norm violations. On the speaker side, since there is no racial meaning to be found "in" C, the speaker of C can truthfully claim that C does not refer to any particular racial group. And on the audience side, the hearers of C can convince themselves (and perhaps others) that their attitude towards x is based on the fact that it is C, not the fact that it is R.

From the point of view of linguistic theorizing, this means that the racial "baggage" associated with code words all belongs in the post-semantics. No race-specific content is encoded in the linguistic meaning of any code word, and code words are identified solely by virtue of having certain generalizable cognitive effects. As we will see below, I think this perspective chops the theoretical terrain too coarsely, and precludes us from marking relevant distinctions between different sorts of terms which effectively prime racial resentment.

Nevertheless, the simple theory nicely accounts for a number of empirical results. For instance, a study by Ismail White (2007) found that harboring anti-black stereotypes was significantly predictive of decreased support for food stamp and Medicaid programs. Yet these stereotypes only had an effect on policy preferences when high-resentment participants were primed with the sentence "food stamps and Medicaid represent important safety nets for many *inner city* families." Interestingly, there was no strong relationship between harboring anti-black stereotypes and policy support when subjects were given the sentence "food stamps and Medicaid represent important safety nets for many *African-American* families." (Scripts substituting the terms *poor* and *American* also revealed no strong relationship between racial biases and policy preferences.)[2]

According to the simple theory, both the term "African-American" and the term "inner city" trigger stereotypes about blacks. Yet only "inner city" offers an ostensibly non-racial interpretive option. Since subjects implicitly believe that basing policy preferences on race-specific grounds would be a violation of the norm of racial equality, they only form preferences based on such grounds when they are *unaware* that they are doing so. The available non-racial interpretation of "inner city" explains why the subjects' racial resentments can be effective in influencing their policy preferences. The term primes them to rely on implicit anti-black stereotypes even though they believe explicitly that these preferences are based on their beliefs and feelings about geographical locations rather than racialized groups.

2. Limitations of the Standard Framing

The standard framing is thus organized around four assumptions.

A. Mendelberg's Assumptions

[2] Hurwitz and Peffley (2005) found strikingly similar effects on support for prison spending. In this study, whites were asked whether they favored prison spending to lock up "violent criminals" in one condition, and "violent inner city criminals" in another. The "inner city" condition resulted in significantly higher support for prison spending.

1. *Normative Assumption:* Americans honor a widespread "norm of racial equality" which separates permissible and impermissible discursive performances concerning race.
2. *Implicit/Explicit Assumption:* Implicit racial appeals are more successful than explicit appeals in mobilizing racial resentment for political ends. When the norm-violating racial content of implicit appeals is made explicit, they become less effective.

Here the truth of the Implicit/Explicit assumption is explained by the truth of the Normative Assumption-- implicit appeals are more effective *because* they are less likely to be seen as norm-violating.

Mendelberg herself does not offer a worked-out theory of the semantics of code words. Yet virtually all extant approaches assume that both of Mendelberg's assumptions are correct. In an attempt to describe the mechanisms of implicit racial communication in more detail, these approaches make two further assumptions about its linguistic vehicles:

B. Semanticist's Assumptions

3. *Semantic assumption:* Code words work by offering multiple competing, plausible interpretive options to their audiences. Some of these interpretive options constitute genuine violations of the norm of racial equality, and others constitute genuine instances of compliance with the norm.
4. *Processing assumption:* Code words permit audiences to grasp norm-violating race-specific messages on an implicit level while consciously assigning semantic interpretations which do not include race-specificity as a component.

This makes the dependency relation between Mendelberg's communication theory and the semantic theory of code words clear. The explanatory power of the semantic theory depends on how generally Mendelberg's assumptions hold. For instance, if Mendelberg's assumptions hold *universally* across all the contexts in which race bears on public discourse, we should expect i) nearly all racial appeals to be implicit, and ii) many of these appeals to be carried by code words. Conversely, if Mendelberg's assumptions are much less generally observed, then there will be fewer "implicit" racial appeals and thus fewer terms that seem to be functioning as "code words."

I will argue that neither of Mendelberg's assumptions have the generality they are usually presumed to have. This section shows that there is no single norm of racial equality that is both i) sufficiently informative to distinguish permissible and impermissible discursive performances involving race and ii) is accepted by all or even the majority of Americans. If my argument is sound, this suggests that many types of racial communication do not neatly fit within Mendelberg's "implicit/explicit" dichotomy. It also suggests that a satisfactory linguistic theory of racial communication should mark interesting differences between the ways in which racialized speech achieves its political goals.

2.1. Examining Mendelberg's Assumptions

We should begin by asking just what content the "norm of racial equality" has within the standard framing. In the absence of a specification of the content of this generally accepted norm, it is difficult to say whether and how a given instance of implicit racial communication allows a speaker or interpreter to circumnavigate those norms. If the norm remains underspecified, this is not only a problem for Mendelberg's framing, but also for Khoo's semantics for code words. For if there *are* no such generally accepted norms, or if there are many *conflicting* norms, then it is difficult to give an informative account of how a term's "coded" effects are explained by the observance of that norm.

To throw the problem into relief, consider a candidate code word C which has proven effective in mobilizing racial resentment. In a society S, there are two plausible specifications of the norm of racial equality, N1 and N2. A race-specific interpretation R of C clearly violates N1. Yet R does not violate N2. All members of the society agree that C can have interpretation R. Yet half of the members of S subscribe to N1, and half subscribe to N2.

If terms like C are widespread, this poses serious problems for the standard framing. The problem for Mendelberg is that for a wide swath of the discourse community of S, C basically has the form of an *explicit* appeal. It is apparent to all members of S that C has interpretation R, but there is still widespread disagreement about whether R is norm-violating. Moreover, the Implicit/Explicit Assumption requires that where some version of the norm of racial equality is dominant, racial communication is effective only when it is *implicit*. Yet if C is indeed an implicit form of racial communication, how can we explain its effectiveness? The problem for Khoo's semantics, on the other hand, is that its commitment to locating racialized meaning in *post*-semantic and often unconsciously drawn inferences and associations leads to an overly general picture of the mechanics of racial communication. As we will see in the next few sections, there is good reason to think that some of the cornerstone terms of mainstream racialized discourse receive explicit, consciously assigned race-specific interpretations "prior" to any inferences that might be computed on those interpretations. Khoo's framework is thus not so much *wrong* as much as it leaves us unable to distinguish between these interestingly different cases.

So what is the content of the norm of racial equality? Recall Mendelberg's statement: "The norm of racial equality is the consensus that the ideology of white supremacy is morally and empirically bankrupt. The norm repudiates the notion that blacks are inalterably *[sic]* inferior and rejects this idea as a justification for treating blacks less favorably than whites." (Mendelberg 2001: 112.) While a norm against white supremacist ideology may be *a* widely shared norm, it is certainly not the norm most relevant to determining which sorts of performances would be perceived to violate the norm of racial equality in *discursive* contexts. While racial slurs and statements such as "blacks are innately less intelligent than whites" might be obvious norm violations, it is far less likely that *possibly* norm-violating statements like "we need to do something about this epidemic of

black-on-black crime," or "blacks are overrepresented on the welfare rolls" violate the norm in virtue of giving voice to white supremacist ideology. [3]

In any case, the standard framing seems to presuppose a far more specific version of the norm of racial equality: a norm against untenable forms of *race-consciousness*. Recall that white Americans' responses to Horton script weakened when Jackson called attention to its racial dimension. One plausible explanation for this shift is that white Americans realized that their racial resentments were illegitimately involved in their support for anti-crime policies. This hypothesis also fits well with the fact that code words like "inner city" work by priming their interpreters to think in race-specific ways, even as they remain unaware that their policy views are based in racial beliefs and preferences. By the same token, a non-coded term like "African-American" is not predictive of low support for welfare because it forces interpreters to confront the possibility that their policy preferences are based in racial preferences.

A norm against "problematic" forms of race-consciousness thus holds that many substantive evaluative judgments should not be based on racial classification.[4] In particular, negative judgments about blacks' perceived lack of commitment to the work ethic, propensity toward criminality, etc, should count as norm violations.

> *Norm against "problematic" race-consciousness*: basing negative judgments about individuals, groups, or policies on their racial "presentation" is prohibited.

On this gloss, "blacks are overrepresented among the poor" will likely seem norm-complying, while "blacks are poor because they are lazy" will be norm-violating.

If there is a universally or generally shared norm of racial equality among American whites, this is likely its content.[5] We have at least good abductive reasons for thinking that the standard framing's Normative Assumption is undergirded by this weak norm against race-consciousness. Recall that Mendelberg holds that "implicit" racial appeals are less likely to be regarded as norm

---

[3] Note that a similar problem applies to Saul's (2017) suggestion that the content of the norm of racial inequality is "don't be racist!" The universality of this norm comes at the cost of its specificity. Campus leftists and rural libertarians might agree that racism is prohibited, but there is no reason to think that these groups will come to any interesting, non-trivial agreements on what sorts of discursive performances count as *racist*. There isn't even very much reason to think that there will be significant agreement *within* these heterogeneous groups. If the norm of racial equality is to have a chance at distinguishing norm-violating from norm-complying performances at any interesting level of generality, "don't be racist!" cannot be its content.

[4] The qualifier "problematic" draws attention to the fact that there may be acceptable forms of race-consciousness.

[5] There is some evidence that something like this norm is central to whites' common conceptions of "racism," which tend to regard many forms of race-consciousness (e.g. pride in non-white racial identities, preferential treatment and "reverse racism" which purportedly victimize whites, etc.) as *ipso facto* morally and politically problematic (Blauner 1992, Kinder and Sanders 1996, Bonilla-Silva 2017.) This is independent evidence that if this is *not* the norm that proponents of the standard framing have in mind, its widespread adoption among many white discourse communities indicates that it ought to be!

violations, and that "explicit" appeals are more likely to be regarded as norm violations. "Explicit racial appeals" are defined as follows:

> By my definition, a racial appeal is explicit if it uses racial nouns or adjectives to endorse white prerogatives, to express anti-black sentiment, to represent racial stereotypes, or to portray a threat from African-Americans. An explicit message uses such words as "blacks," "race," or "racial" to express anti-black sentiment or to make racially stereotypical or derogatory statements. (Mendelberg 2001: 8)[6]

In practice, Mendelberg's conception of explicit racial appeals is broader than this definition. In some cases, mere *reference* to a racial category is sufficient for making an explicit racial appeal. For instance, the following quotation from North Carolina senator Jesse Helms is treated as an explicitly racial appeal:

> The big factor in this election will be whether there will be a balance to the efforts of Jesse Jackson, who came into this state earlier this year to meet with Governor Hunt and then announced that he was going to register, I-forget-what-it-was, 200 or 300-thousand *blacks* for the sole purpose of defeating Jesse Helms (Mendelberg 2001: 101).

In another instance, the following sentences appear in the "explicitly racial" condition of an experimental script of Mendelberg's own design: "In his last campaign, Hayes said that some people, *especially blacks*, take advantage of welfare at the expense of hard-working taxpayers. He claimed welfare had become a way of life for many *black people*…" (Mendelberg 2001: 205). This equation of the notion of a "norm-violating" performance with a "race-specific" performance is echoed by Khoo's simple theory, which holds that code words work by saying something ostensibly "non-racial" while leading the audience to form beliefs which are "racial" in character.

The set of performances which the standard framing deems to be "norm-violating" is thus rather heterogeneous. If anything from verbal references to race to explicit endorsements of white supremacist ideology belongs in the class of explicitly racial appeals, then a weak norm prohibiting negative forms of race-consciousness seems the most natural way to section off all of the most likely

[6] There are already some minor reasons to be suspicious of this conception of explicit racial appeals. One problem is that again, outside of a handful of obvious examples, it is not at all clear what it means to make a claim involving terms like "blacks," "race," or "racial" in order to "endorse white prerogatives," "represent racial stereotypes," and the like. Presumably, features of the context (who the speaker is, what the surrounding discursive context is, etc.) will matter in determining whether it is-- which is just to say that it's unlikely that the distinction between implicit and explicit appeals lies solely with a difference in the sort of lexical items each sort of appeal contains.

performances that will be taken to count as norm violations.[7] It also provides the most natural explanation of why implicit appeals lose their force when their racial dimension is made explicit.

The problem is that weak norms against race-consciousness aren't as widespread as the standard framing assumes. In one of the largest studies to date (6300 respondents) on the impact of implicit and explicit racial appeals on views of public policy, Huber and Lapinski (2006, 2008) found that while white subjects of all racial resentment levels *were* more likely to regard explicit racial appeals as illegitimate, explicit appeals were *just as likely* as implicit appeals to affect respondents' support for strengthened welfare work requirements, affirmative action, or government aid to blacks. Moreover, lower-education whites did *not* treat explicit appeals as less legitimate than implicit appeals, and were equally susceptible to being primed by both implicit and explicit appeals. (The explicit appeals in question contained scripts which made stereotype-consistent verbal references to "blacks," while the implicit appeals contained no such verbal references, but did contain race-specific images.)

There is also evidence that many whites tolerate explicitly racial appeals *even* when those appeals are clearly associated with membership in unambiguously racist groups. In a study on white support for Confederate flag symbols in Georgia, Hutchings et al (2010) discovered that support for the Confederate flag was only moderately decreased when subjects were exposed to news stories about Klansmen and neo-Nazis who defended the value of the flag on explicitly white supremacist grounds. The authors suggest that "explicit racial appeals do not necessarily disturb all whites, even when they clearly represent a violation of the norm of racial equality" (Ibid. 1181.)

Finally, Huber and Lapinski point out that educational attainment seems to play a role in determining what sorts of racial discursive norms are honored within particular sub-communities of whites:

> [Mendelberg's model] presupposes that all respondents recognize and reject explicit appeals as illegitimate because they violate widely held norms against racial discrimination. But rejecting an explicit appeal as illegitimate requires attachment to the egalitarian norm, recognition that the message is violating this norm, and the ability to reject a message that violates the norm. All three steps of this process are more likely among those with higher levels of education.
> (Huber and Lapinski 2006: 433)

[7] Of course, if could be objected that the "mere verbal references" above are instances of portraying a threat from African-Americans, or exploiting anti-black stereotypes. I myself am sympathetic with that view, but this could be a point of reasonable disagreement. The fact that these statements do not strike some as *obviously* racially inegalitarian indicates that the message may be more or less "explicit" to different interpreters. In any case, the question that matters is whether *white Americans in general* see these appeals as norm violations, whether they reject them as illegitimate sorts of appeals, and whether the appeals themselves mobilize white racial resentment. As we will see, "explicit" appeals indeed do mobilize racial resentment effectively. Yet such appeals are not always seen as norm violations, much less rejected on the basis of norm-violating content.

This has provocative implications for the generality of Mendelberg's own data concerning racial norms. Much of her experimental work in defense of the Implicit/Explicit assumption relies on data collected from about 200 Michigan respondents who were low in racial resentment (only 16% coded as having "high resentment") and highly educated (about 46% had completed at least some *post-graduate* education.) (Huber and Lapinski 2006: 436) This suggests that Mendelberg's conclusions may be based on unrepresentative samples.

All of this presents a *prima facie* problem for Mendelberg's theory. The norm of racial equality is supposed to explain why implicit appeals are more effective than explicit appeals, for instance, as well as what separates norm-complying and norm-violating discursive moves. Yet specifying the content of this norm renders it too weak to distinguish norm-violating and norm-complying performances, or else leaves us with a norm which seems not to be widely shared across white American discourse communities. We should thus conclude that the Normative Assumption is false. There is not one norm of racial equality, but *many* possible specifications of the norm which are operative in different discourse communities.

Of course, Mendelberg may respond that the problem is only apparent. It still may be the case that white Americans are generally hesitant to appear racist. The differences between white American subgroups may simply lie in *how strongly* the norm of racial equality binds social interactions within each subgroup.[8] Imagine two subgroups of white Americans, A and B. In subgroup A, the norm is very weakly observed. Members of A are undeterred by the possibility that they will face censure for their race-related thoughts and words, *but only* when they can be reasonably sure that their norm violations will remain "in-house," as it were. When they do not have such assurance, A-members modify their conduct with respect to the norm so as not to appear racist. In subgroup B, the norm is also very weakly observed. Yet unlike A-members, B-members *simply do not care* if they face censure for their race-related thoughts and words. They only observe their "in-house" norm, and never modify their conduct in relation to others. We should expect A-members to be swayed by explicit racial communication only when they feel they can get away with it, and B-members to be swayed by it more often than that.

Now, on the plausible assumption that most whites are closer to being A-members than B-members, we have a ready explanation for findings that suggest that whites are no less responsive to explicit messages than they are to implicit ones. Perhaps lab settings do not provide many subjects with sufficiently strong reasons *not* to respond to explicit racial communication. But that's no evidence against Mendelberg's basic ideas that i) most Americans are constrained by at least some form of the norm of racial equality, and ii) that we should generally predict that implicit racial communication will on balance be more effective than explicit communication.

We should grant Mendelberg that there are many ways of reading the data here, and many possible stories to tell about why "norms of racial equality" are not always seen as binding.[9]

---

[8] Thanks to an anonymous reviewer for raising this possibility.

[9] One possibility is that the recent mainstreaming of right-wing populism (discussed in the next section) incentivizes *defection* from the norm of racial equality among people who were otherwise committed to it.

Regardless of which particular story we endorse, however, notice that the explanatory generality of the basic theory is still hindered. Interpreters often honor a distinction between i) whether a performance violates a generally accepted norm of racial equality in the broader discursive community and ii) whether a performance counts as *racist* according to the standards they themselves endorse. Those who rely on explicit racial appeals are generally aware that their speech violates *some* racial norms, and perhaps even violates *dominant* racial norms. Yet when these norms are seen as having an origin "outside" the relevant discursive and practical communities, they are not seen as binding. Moreover, it seems that *quite a lot* of people participate in and respond to explicit racial communication, even when they are at risk of facing substantial censure, and that the number seems to be increasing-- in other words, there might now be more B-members among white Americans than Mendelberg supposes. What we are still lacking is an account of racial political speech that makes sense of *these* sorts of norm violators, regardless of "where they came from."[10]

To be fair to Mendelberg, the tenor of public political discourse in the United States has changed significantly since the publication of her landmark work in 2001. Her theory remains explanatorily adequate for many forms of implicit racial communication. Yet these significant shifts in public discourse do suggest that Americans like the one Mendelberg quotes in the wake of the Horton strategy can no longer be written off as mere statistical outliers: ""The average voter… just plain don't feel guilty for being scared of black criminals… They didn't understand why it was racist to talk about reality"" (Mendelberg 2001: 171). In the next section we'll canvas one of the central rhetorical strategies used to target precisely these sorts of voters.

3. Candid Racial Communication

The question now is then *how do* racial appeals work in contexts where substantive norms against "problematic" forms of race-consciousness are not operative? This section reviews two examples of candid racial communication. Candid racial communication (CRC) differs from implicit racial communication in a number of respects. First, CRC uses openly race-specific language as a way of licensing its audience to think and speak in race-specific ways. Second, CRC often achieves its communicative goals by making race *more* rather than less salient to its interpreters. Finally, CRC has a *performative* dimension. Its role is not only to target those who do not conform to certain norms, but also to *create* a community of "norm-violators."

Here the hypothesis would be that the norm of racial equality is a kind of "default" as long as sanctions for defection are leveraged in one's discourse community. Recent racial appeals seem to erode this sanctioning behavior, and so disincentivize commitment to the norm of racial equality. A story about the rise of right-wing populism should thus see this politics as not only *reflecting* the pre-existing racial attitudes of non-dominant discourse communities, but also as *creating* new discourse communities in which norm-defection is not sanctioned. Thanks to Rachel McKinney for pointing this out.

[10] One might argue that Mendelberg's theory shouldn't be expected to explain something it is not designed to explain. That may very well be the case, but given that the standard framing is still widely treated as the only game in town for making sense of racial appeals in political communication, it does make it seem that we need a new game.

3.1 Two Examples of Candid Racial Communication[11]

#1: Gingrich at the debate

Stanley (2015) recounts the following interaction:

> …[In] a debate during the Republican primary presidential campaign in 2012, Juan Williams asked a candidate, Newt Gingrich:
>
> You recently said, "*black Americans should demand jobs, not food stamps*. You also said, "*poor kids lack a strong work ethic*," and proposed having them work as janitors in their schools. Can't you see that this is viewed, at a minimum, as insulting to all Americans, but particularly to black Americans?
>
> Gingrich answered, "No. I don't see that," and received a loud ovation from the audience. (Stanley 2015: 157)

#2: Donald Trump's "Law and Order" speech, 2017

In July 2017, Trump gave a speech to a large audience of law enforcement officers on Long Island. Throughout, Trump warned of the dangers posed by lax immigration laws and decried gang violence perpetrated by non-American "animals."

> I have a simple message today for every *gang member and criminal alien* that are threatening, so violently, our people. We will find you, we will arrest you, we will jail you, and we will deport you.

He went so far as to encourage law enforcement officers to "get rough" with suspects:

> When you see these *thugs* being thrown into the back of a paddy wagon, I say "please don't be too nice." Like when you guys put somebody in the car, when you're protecting their head… I said, you can take the hand away, ok?

Finally, he promised to protect the United States against terrorist activity:

> We're also working… on a series of measures to keep our country safe from *crime and terrorism*. And in particular, *radical Islamic terrorism*. A term never uttered by the past administration. Did anybody ever hear that term? I don't think so. But you heard it from me. (Dice 2017)

[11] All emphases in the quotations are mine.

Each of these discourses 1) carries an evaluative *race-specific* message, 2) conspicuously lacks any discursive move to hide the race-specificity of the appeal,[12] and 3) includes a performative dimension which *licenses* its audience to assign race-specific interpretations to what is being said.

Gingrich's comments, which include the phrase "black Americans," do not even allow his interpreters (or himself) to plausibly deny that these comments were not racial in character. Yet clearly this did not alienate his (conservative) audience. Presumably, the people loudly applauding for Gingrich's refusal to back down from the claim that "black Americans ought to demand jobs, not food stamps" did not think that they were thereby being racist (although they likely would have happily granted that they were not being "politically correct.") And virtually all of Trump's second law and order speech oscillates between praise for American law enforcement and dire threats about the threats that "illegals," "terrorists," and "thugs" pose to the United States. It is clear that the audience would be missing a key point of the discourse if they failed to think of the members of these categories in race-specific terms. After all, the problem isn't *terrorism*, it's *radical Islamic terrorism*.

3.2 The normative structure of candid racial communication

These discourses illuminate a core difference between the normative structure of implicit racial communication and that of CRC. While implicit racial communication thrives in contexts where everyone generally acts as if they are constrained by a generally accepted norm of racial equality, CRC thrives in contexts where people *don't* act as if they are constrained by that norm. Here I consider two reasons why such a norm might not be seen as binding.

First, perhaps candid racial communicators simply presuppose that their audiences accept a *different* norm, or at least acceptance of a meta-norm which treats the "standard" norms against race-consciousness as more or less defeasible. Here an interpreter need not think that such norms are unimportant; she need only think that other considerations sometimes outweigh those norms. For instance, the applause that greeted Gingrich's refusal to back down from the claim that "black Americans should demand jobs, not food stamps" might show that his audience took his statement to be justified by its underlying commitment to a deeper meritocratic norm. Part of the power of such norms is that they at least appear to be racially egalitarian-- hard work is valuable and dependency is disvaluable, no matter what race you are. Yet Gingrich's appeal takes the form of an appeal to common sense: if we're going to be honest with ourselves, it's *black* Americans in particular who could do better to work harder and complain less. In a similar vein, Trump's speech can be seen as giving his audience a rational and moral imperative to recognize the imminent threat that certain agents impose to American well-being. Here, the disvalue of racism is outweighed by the

[12] This feature separates true cases of candid racial communication from what Saul (2017) has identified as "racial figleaves." Racial figleaves make racist propositions seem more reasonable by introducing explicit "hedging" moves designed to lessen the impact of the objectionably racist content. A common figleaf is the well-known "I'm not racist, but…" followed by an explicitly racist assertion. On Saul's approach, figleaving indicates that some sort of norm of racial equality is in effect, since the need for the figleaf would be unclear otherwise. My analysis of candid racial communication is not in tension with Saul's account of "figleaves." We are simply interested in different phenomena.

benefits of safety. Latino "gang members," "criminal aliens," and "thugs" are what we must be on guard against, and, despite what those more attached to political correctness would have you believe, we must be on guard against "radical Islamic terrorism… you heard it from me."

Of course, signaling the legitimacy of a (meta-)norm which trumps the norm of racial equality leaves the candid racial communicator vulnerable to the charge of racism in a way that the implicit racial communicator is not. When such cases arise, the implicit communicator is forced to defend herself by demonstrating that she does not violate the common racial norms that she (ostensibly) accepts. Candid racial communicators, on the other hand, usually find themselves in the position of arguing that some other concern trumps the norm of racial equality. One extreme form of this strategy is to signal that the norm of racial equality is *actively harmful* to the interests of his audience.

It is worth treating this point in a bit more detail. Candid racial communication is just one of a broader suite of rhetorical strategies associated with right wing populism. As the linguist Ruth Wodak (2015, 2017) has pointed out, such strategies depend on a "dichotomous view of society (a merger of anti-elitism with a nativistic nationalistic anti-pluralism)" which is disseminated "by continuously creating or maintaining confrontations with those who are seen as not being part of the 'real America,' the 'real France,' 'the real Austria,' and so forth" (Wodak 2017). Within the discourse communities in which these strategies reign, white Americans are powerfully incentivized to think of themselves as "the people," denizens of the "real" America, which is under constant threat and attack from various enemies. In particular, the people must be on guard against the "elites" and the "establishment," which consolidate power through social and political gambits designed to take status, wealth, and power away from that "real" America and redistribute it among "minorities" and "immigrants."

Both in the United States and in other western democratic states, the acceleration of globalization, the financialization of the global economy, resurgence of the fundamental rifts that animated the culture wars of the 1980's and 1990's, and increasing class inequality have all done their part to intensify white resentment, to the extent that political elites no longer shy away from embracing right-populist discourse. Candid communication thus particularly thrives upon and promotes an identity politics heavily invested in a dichotomy between those who "belong" (and are thus deserving of recognition, respect, and resources) and those who don't (and thus are not). Evidence from social psychology suggests that right-populist agendas can only benefit from raising the specter of a progressive "minoritization" of the "home" nation. For instance, a study by Maureen Craig and Jennifer Richeson (2014a) found that whites espouse more conservative views on both racial and non-racial issues when informed that America is fast becoming a majority non-white nation. In another study, they found that whites are more likely to exhibit both implicit and explicit forms of bias against blacks, Latinos, and Asian-Americans when informed that these groups are becoming increasingly represented in their neighborhoods (Craig and Richeson 2014b.) Sometimes it is merely the increased visibility and representation that non-white groups have achieved that make these groups seem intrinsically threatening to whites. For instance, a 2017 poll

found that 45% of whites who voted for Donald Trump believe that whites are the racial group which faces the *most* discrimination in America. This belief is in turn highly predictive of right-wing voters' increased tolerance for neo-Nazi and white supremacist groups (Saletan 2017).

The overarching political function of candid communication is thus to make vivid a struggle between a deserving "us," and an undeserving, (often non-white) "them." The endgame of right-populist rhetoric is to invite an audience into the fold of the "real America," and to rally the troops for a confrontation against those outside the fold, particularly "the elites."[13] And in many cases, *anti-racist norms themselves* are seen as tools of the elite. The performative dimension of candid racial communication here is to create a discursive context in which the speaker *is not bound* by charges of racism that are issued from outside the community she is addressing. To these audiences, the question of whether they are observing the norms of "politically correct" racial propriety in responding to not-so-coded racial messages is far less important than whether these responses demonstrate their inclusion within the group candid racial communication is designed to appeal to. It thus becomes clear precisely who "they" are in the epigraph which opens this paper, and precisely why one should be "honored" to be excoriated by "them": "Let them call you racists. Let them call you xenophobes. Let them call you nativists. Wear it as a badge of honor."

The practice of candid racial communication can thus represent the norm of racial equality as highly defeasible, or as actively harmful to an audience's conception of their social and political ends. A candid racial communicator might even toggle between these strategies in order to enhance the general appeal of her message to different sorts of audiences. Yet in either case, what candid racial communication strives to do is create a discursive context-- or even a discursive community-- in which anti-racist norms are seen as not binding. As we will see, I think that this feature of candid racial communication gives us occasion to draw some interesting conclusions about the relationship between discourse norms and linguistic meaning.

4. Racialized Terms and Referential Modification

We have seen that the norms governing racial communication can differ widely across American discourse communities. This normative variance allows for a diversity of rhetorical strategies engineered to target and weaponize voters' racial resentment. In the remainder of the paper, we turn to the question of what upshots this has for linguistic theorizing about the particular expressions which carry racial appeals.

Here is what seems clear: in contexts where a norm against problematic race-consciousness is in effect, code words are the coin of the realm in exploiting racial resentment. The fact that interpreters can easily convince themselves that they are not violating such norms in using and

[13] Former Trump administration chief strategist and former executive chairman of Breitbart News Steve Bannon has been notably forthcoming about this goal, and is currently involved in a self-described campaign to "declare war" on the "Republican establishment," which is responsible for "economic hate crimes" against the "American people": "You have to have a sense of urgency. Nobody is safe. We are coming after all of them and we are going to win." (Schwartz 2017).

responding to these code words indicates that the semantic link between these terms and racialized meanings is rather indirect. Since it does not appear that we need resources stronger than Khoo's simple theory in figuring out how they work, I grant that what we might call "true code words" trigger post-semantic inferences or associations which get across racialized content in the way Khoo outlines.

Yet in contexts where norms against race-consciousness are not operative, we observe more direct semantic relationships between racialized terms and race-specific meanings. Our examples of candid racial communication indicate that there are terms which i) are *directly predicated* of particular racialized groups and ii) elicit racial resentment by making salient *race-specific* interpretive options (as opposed to true code words, which function by making salient race-neutral interpretive options.) This is an important distinction that Khoo's theory does not mark.

Strictly speaking, there are two classes of terms that bear this more direct relationship to race-specific meanings. First, there are what we might call *taxonomic racialized terms*. These are terms whose extensions include all and only individuals that are racialized in a particular way-- Gingrich's "black Americans" and Trump's "radical Islamic terrorism" are obvious examples. These terms are directly predicated of the racialized groups that correspond to them. Second, there are what we might call *referentially modified racialized terms*. These are terms with intuitively non-racialized meanings-- "poor kids," "gang member," "criminal alien," "thug"-- which are modified in context to include particular racialized groups-- in this case, *black kids* and *Latino immigrants*. In each case, these terms allow the audience to become explicitly aware that particular racialized groups are under discussion. But if these are code words, they are code words that are in no sense coded. I thus propose to drop talk of "code words" altogether in what follows, in favor of "racialized terms."

While we have seen that there are interesting discussions about what sorts of background norms make the *use* of taxonomic racialized terms like "radical Islamic terrorism" or "black Americans" more or less permissible, there is nothing particularly special about taxonomic racialized terms in light of our present purposes. For one thing, we should expect their semantics to behave no differently from referring expressions like "wild horse" or "yellow daffodil." For another, taxonomic racialized terms are pretty clearly not the sorts of things that Khoo's account is designed to model. I thus largely drop discussion of them in what follows.

Far more interesting is the class of referentially modified racialized terms. These terms seem to acquire a race-specific referential content[14] in certain contexts. Given a background acceptance of certain norms for referring to racialized groups, these terms function less like "code" for racialized groups and more like *explicit descriptors* for them. If we can make good on this hypothesis, we have identified a class of expressions that puts pressure on Khoo's assertion that the simple theory "may be all we need to understand coded speech" (Khoo 2017: 32).[15]

---

[14] As we will see in Section 4.3, this is importantly different from having a race-specific "conventional meaning" at the context.

[15] This is on the assumption that "a theory of coded speech" needs to account for the referentially restricted racialized terms that are of interest to me: "terrorist," "immigrant," "thug," "gang member,"

Referentially modified racialized terms might undergo contextual processes of *referential restriction* or *referential expansion*. We characterize each in turn.

4.1. Referential Restriction

The term "referential restriction" is borrowed from Justina Diaz Legaspe (2018). According to one orthodoxy in the literature on slurs, slurring terms generally have non-derogatory "neutral counterparts." On some views, it is this neutral counterpart which provides the thin referential content of the slur. So, for instance, the referential content of the derogatory term "mick" is often taken to be identical to that of the non-derogatory expression "Irish-American." On this view, "neutral counterparts of slurs provide [slurs] with the correct application criteria" (Diaz Legaspe 2018: 234). Against this orthodoxy, Diaz Legaspe argues that in many cases, slurs and their neutral counterparts do not co-refer. One piece of evidence for this claim is that slurs are often used to pick out a *subset* of members of the relevant "neutral counterpart" class. So, to adapt a famous line of Chris Rock's, consider:

1) I love Irish-Americans, but I hate micks.

If the neutral counterpart *Irish-American* provides the correct application criteria of "mick," and if "Irish-American" and "mick" bear the same referential content, 1) expresses a contradiction. The fact that 1) does not strike us as a contradiction is that we naturally interpret the use of "micks" as referring to some *subset* of Irish-Americans, namely those the speaker deems worthy of contempt.[16] This is a referentially restricted use of the slur.

I am not suggesting that we understand racialized terms as slurs *per se*. Yet there is an interesting parallel between slurs and racialized terms. Intuitively, a term like "terrorist" has a neutral referential content like *perpetrator of ideologically motivated violence*. Yet patterned uses of "terrorist" indicate that this term is often referentially restricted in context to something like *Muslim perpetrator of ideologically motivated violence*.[17] Consider, for instance, the fact that right-wing violent white extremists

---

"criminal alien," etc. But even if a theory of coded speech need *not* account for these terms, the relative scarcity of proposals concerning how we *ought* to think about these terms makes my positive proposal of independent interest.

[16] An anonymous reviewer suggests that 1) strikes us as felicitous because "Irish-American" and "micks" pick out generic classes which share members but are not entirely co-extensive. This is a substantive hypothesis about the *mechanisms* behind referential restriction in the case of slurs, and one that may well be true. However, I do not take a stand on the issue here.

[17] As an anonymous reviewer points out, this commits me to the claim that referentially restricted readings of "terrorist" entail the truth of the sentence "all terrorists are Muslims" at certain contexts. This may seem like an implausible consequence of my view. For instance, if we occupy a context where the term "dogs" is conventionally restricted to *yappy dogs*, we are not tempted to think that "all dogs are yappy" is a true sentence, all things considered. So why think that "all terrorists are Muslims" is ever true in context?

I think that this consequence seems less implausible when we consider two things. First, when entailments from referentially modified racialized terms to their corresponding universal quantifications do go through, it is not by virtue of the *conventional meaning* of racialized terms alone. Rather, it is by virtue of referential modifications that these terms undergo in context (more on this in Section 5). These contexts, in turn, must meet rather specific conditions in order for these terms to actually undergo those modifications.

are rarely labelled "terrorists," despite their being no less ideologically motivated than their Muslim counterparts (Yin 2013). Or consider a real-life example of a referentially restricted use of "gun-toting drug dealer." In context this term only seems to pick out a subset of armed drug dealers (namely urban non-whites)[18]:

> I had as [assistant U.S. attorney who] wanted to drop the gun charge against the defendant [in a case in which] there were no extenuating circumstances. I asked, "Why do you want to drop the gun offense?" And he said, "He's a rural guy and grew up on a farm. The gun he had with him was a rifle. He's a good ol' boy, and all good ol' boys have rifles, and it's not like he was a gun-toting drug dealer." But he was a gun-toting drug dealer, exactly. (quoted in Alexander 2012: 118)

Finally, consider an example of referential restriction in the case of "immigrant," due to Jennifer Saul (Saul 2015):

> I am an immigrant. I came to the UK 20 years ago from the US to teach philosophy at the University of Sheffield, where I am now a professor. My American accent remains very strong. I used to be surprised when, despite hearing me speak, people would express anti-immigration sentiments to me, with a clear expectation of agreement. I would tell them that I am an immigrant. "I don't mean *you",* they'd respond, surprised that I count myself as an immigrant. This shows that seemingly neutral words – like "immigrant" – are not always used in a neutral way. ["Immigrant" and "migrant"] is increasingly used by the media to describe the large numbers of desperate people travelling into and across Europe, fleeing war and persecution. (Saul 2015)

Second, the notion that "all terrorists are Muslims" is true at some contexts seems less implausible when we consider that categorizational practices involving social kinds often shape the nature of those kinds. Unlike the term "dogs," changes in the use of words like "terrorist" can engender changes in the nature of the *objects* that these terms purport to pick out. Even if a linguistic community were to start using the term "dog" to pick out, say, phonebooks, *dogs* would remain dogs. Since dogs are an "indifferent kind" (Hacking 1999), members of this kind do not modify their behavior or natures in response to how they are labelled. "All dogs are yappy" is false because regardless of how we restrict the extension of "dog," *as a matter of fact* some dogs are yappy, and some are not. Contrast this with "terrorist." "Terrorist" is a contested term whose extension is often "up for grabs" among different discourse communities. Were a discourse community to systematically change its practices of applying the term "terrorist," the terrorists themselves would change, too, in a way that reflected that categorizational practice. In this way, what terrorists *are* is "up to us" in a way that other bits are the world are not (those who engage in political and metalinguistic contestations about who the "thugs," "terrorists," "illegals," "gun-toting drug dealers," "immigrants," and "criminal aliens" are, if they are anyone at all, recognize this.) My claim is that where a particular negotiated modification of a racialized term is dominant, that term may carry entailments at that context that it lacks at other contexts, and that what those terms entail may be genuinely *true* at some contexts while being false at others (Muhlebach 2019 seems to develop this idea along inferentialist lines.) Section 5 will be concerned to show that this hypothesis, baroque as it may seem, does not require us to abandon any standard commitments about the nature of conventional meaning.

[18] Thanks to an anonymous reviewer for drawing my attention to this example.

The semantic value of "immigrant" that is analogous to a slur's "neutral counterpart" is something like *person who comes to live permanently in a foreign country.* As an American living in the UK, Saul clearly belongs in extension of "immigrant." Yet to her interlocutors, "immigrant" secures reference to some subset of the class of people of foreign origin who have taken up residence in the UK. Saul suspects that at these contexts, "immigrant" is being used in a way that picks out groups who are specifically racialized, particularly those of Middle Eastern, Asian, or African origin.

These examples of referential restriction allow us to see that in certain contexts, terms can receive semantic values that differ from those we would expect to be associated with their conventional meanings. The following sentences, for instance, assert controversial views, but there is an intuitive sense in which they are felicitous:

2) Dylann Roof is a perpetrator of ideologically motivated violence, but he is not a terrorist.
3) Bill is a rifle owner who sells illegal drugs, but he is not a gun-toting drug dealer.
4) Jenny Saul is a foreign-born permanent resident of the UK, but she is not an immigrant.

These sentences indicate that the uses of "terrorist," "gun-toting drug dealer," and "immigrant" pick out some subset of the individuals in the "neutral" extension of these terms, and that Dylann Roof, Bill, and Saul do not meet the requirements for membership in that subset.

Of course, there is also a reading on which these sentences seem infelicitous. (How could Dylann Roof *not* be a terrorist? How could any rifle-owning drug dealer *not* be a gun-toting drug dealer? How could any foreign-born permanent resident not be an immigrant?) Yet what matters for my purposes is that there are some contexts--perhaps a relatively small number of them-- in which 2)-4) are genuinely felicitous expressions. Indeed, while the non-racialized, infelicitous reading may be the most "natural" reading, it is possible to access the racialized felicitous reading of sentences like 2)-4) by occupying or even imagining oneself in the relevant discourse contexts where a referentially restricted interpretation of terms like "terrorist," "immigrant," "thug," "gun-toting drug dealer," or "criminal alien" reigns--indeed, most of us traverse such contexts daily. This indicates that referential modification is an important mechanism by which terms become infused with race-specific meanings within particular discourse communities. We return to this point in the next section.

4.2 Referential expansion

Referentially restricted racialized terms selectively pick out racialized members of an otherwise race-neutral class. Yet in some cases, candid racial communicators might *expand* the referential content of a race-neutral term in order to include racialized groups. Consider a recent example. In May 2018, Donald Trump referred to certain immigrant groups as "animals" in a roundtable discussion of immigration policy:

> We have people coming into the country, or trying to come in — and we're stopping a lot of them — but we're taking people out of the country. You wouldn't believe how bad *these people* are. *These aren't people. These are animals.* And we're taking them out of the country at a level and at a rate that's never happened before. And because of the weak laws, they come in fast, we get them, we release them, we get them again, we bring them out. (Lind 2018, emphasis mine)

In the wake of these explicitly dehumanizing comments,[19] substantial disagreement arose concerning just what "animals" referred to. Trump's detractors claimed that "animals" was intended to refer to immigrants from Central and South America quite generally, while defenders argued that Trump was referring to that subset of immigrants responsible for violent crimes. The White House even released an official memo claiming that "animals" denoted members of the gang MS-13, whose membership is primarily of Central American origin. Despite this referential indeterminacy, the term "animals" was quickly adopted by many of Trump's supporters:

> Trump doubled down on controversial comments he'd made earlier this month about the drug gang. *"What was the name?" the president prodded the crowd, who yelled back "Animals!" "They're not human beings," Trump added,* saying that they use "glaring loopholes in our immigration laws" in order "to infiltrate our country" and rape, murder and "cut people up into little pieces." (Taylor 2018, emphasis mine)

As our characterization of candid racial communication should lead us to predict, Trump's audience does not deny that the referent of "animals" is racialized in more-or-less specific ways. Moreover, Trump's use of "animals" effectively gives his audience a more or less conventionally shared way of talking about the immigrants traversing the southern border of the United States. Sustained by common beliefs about what the term refers to, "animals" takes on the semantic potential to refer to racialized groups in a way that it lacked previously. The conventional, socially coordinated awareness of this re-purposing of the term is of course illustrated to great effect in the example above by one of those most intimate forms of discursive synchronization, the "call and response."

5. Community-specific speech, and an advantage of the view

I have argued that the structure of the racial norms within a discourse community can affect the semantic values of the terms that are used to pick out and evaluate racialized individuals, groups, and conduct within those communities. These constraints have a conventional flavor: everyone in the relevant community knows that certain terms ("thug," "criminal alien," "terrorist," "animal") pick out more or less specifically racialized groups. Outside these communities, these terms don't have this same semantic potential.

[19] While not explored in this paper, the practice of referentially expanding terms for non-human animals to incorporate particular racialized and ethnicized groups is on display in the dehumanizing tactics reflected in propagandistic hate speech. See Smith (2011), Tirrell (2012), Maynard and Benesch (2016), and Marques (2019).

To this extent, rather than trying to assimilate referentially modified racialized terms to certain types of code words or slurring expressions, we ought to cast them as specific types of "community-specific speech" (Hebert and Kukla 2016.)[20] On Hebert and Kukla's (2016) use, community-specific speech, including jargon, slang, and inside jokes, is speech that is "used by and directed at community insiders" in a way that allows members of the relevant community to identify fellow insiders and to consolidate group membership. If racialized terms in the context of candid racial communication are a type of community-specific speech, then the relationship between the proprieties of use for a racialized term R in a context and R's referential content are more closely related than is usually supposed.[21]

Yet it is important to distinguish the thesis that racialized terms can have more than one *referential content* from the idea that racialized terms can have more than one *conventional meaning*. The conventional meaning of an expression E is what a user of E "grasps" when she uses E correctly. This conventional meaning is determined by the conventional rules of the language, and is widely taken to play a key role in determining the referential content of E. Yet not all conventional meanings are created alike. Some terms-- notably proper names like "Aristotle"-- have conventional meanings which allow them to pick out the same individual in all contexts. At the other end of the spectrum, terms like "I," "today," "tomorrow," "that," and "here" have conventional meanings which allow them to pick out *lots* of different individuals depending on features of the context in which these expressions are used. For instance, the "I" in "I am hungry" picks out *Bismarck* when *he* utters it, but "I" picks out *Keisha* when she utters it, "today" picks out *Wednesday* only if it is uttered on a Wednesday, etc. The *referential content* of E can thus be understood as whatever individual E picks out at a context. Of course, sometimes a term might have a perfectly intelligible conventional meaning while having *no* referential content (e.g. "unicorn," "Santa Claus," "witch," etc.)

Now, what I have called "racialized terms" sometimes pick out racialized groups, and sometimes do not. Whether they do so depends on features of the context, and in particular the observed racial norms at the context. But note that explaining this variation in terms of referential content does not inherit many of the problems that theorists have confronted in accounting for the semantics of code words. On one traditional set of assumptions, conventional meanings are fairly "brittle" features of linguistic items. But code words are variable and plastic. The brittle conventional meanings that these terms express could not account for all that variation if they tried, and so meaning variation must be explained by something other than conventional meanings (usually by "post-semantics" or "pragmatics.")

Consider a recent account which demonstrates the difficulty of reconciling the plasticity of racialized terms with the received view of conventional meaning. Like my account, Jason Stanley's

---

[20] Hebert and Kukla are interested in the pragmatic dimensions of such speech. My argument makes the further claim--which the authors may or may not endorse-- that the "community-specificity" of the speech itself can affect the referential content of the terms used in the discourse.

[21] This basic view of the relationship between proprieties of use and the semantic content of racialized vocabulary is also endorsed by Deborah Muhlebach, who develops the view within an inferentialist (and therefore non-referentialist) semantic framework (Muhlebach 2019.)

(2015) recent approach to the semantics of code words recognizes that some code words can transcend their "coded" character. Through patterns of "repeated association" between terms and race-specific images, race-specificity becomes "part of the meaning" of those terms, and perhaps even part of the "conventional meaning" of such terms:

> [When] propagandists use repeated association between words and images, they are forming connections that serve as the basis of conventional meaning...When the news media connects images of urban Blacks repeatedly with mentions of the term "welfare," the term "welfare" comes to have the not-at-issue content that Blacks are lazy. At some point, the repeated associations are part of the meaning… (Stanley 2015: 138)

Racial meanings are thus not located in the *referential* content of the code word, but in the "conventional meaning" of the term more broadly. Stanley's attempt to cash out this proposal relies on the distinction between (non-racial) *at-issue content* and (racial) *not-at-issue content*. Roughly, an utterance's at-issue content is its main point or contribution, while not-at-issue content includes more ancillary information. Consider:

5) I rode my bike to the park in the rain.

5) tells you i) that I own a bike, ii) that I know how to ride it, iii) that the park is in reasonable proximity to wherever my bike journey began, iv) that it was raining, etc. Yet only some of this information is central to the topic of discussion, and other information is peripheral. The main point-- the at-issue content-- of 5) is to inform you that *I traveled from one location to another on a bike.* Anything else that the utterance might convey-- that I own a bike, that it was raining, etc.-- belongs to its not-at-issue content.

A more technical way of distinguishing at-issue and not-at-issue content is to consider what elements of an utterance are available for direct denial. Consider what would happen if you were to negate 6) by use of 7):

6) I rode my bike to the park in the rain.
7) No you didn't!

Naturally, I would understand you to be saying that *it is not the case* that I rode my bike to the park. Yet 7) does not directly target various other components of 6). For instance, it's not plausible that your utterance as it stands successfully denies that I have a bike, or that it was raining. In order to make that sort of move, you would have to refocus our conversational attention on some part of my claim that I was assuming that you would take for granted. This requires linguistic moves more sophisticated than direct denial, including:

8) Wait, it didn't rain today!
9) I didn't know you had a bike!
10) The park is way too far for you to ride your bike to it!

Stanley leverages this distinction to account for the meaning of coded words. Code words like "inner city" and "welfare" have reasonably well-defined at-issue contents referring to *urban cores* and *public assistance*, respectively. Yet these also carry race-specific not-at-issue contents concerning *poor blacks* and *lazy blacks*.

This theory nicely accounts for at least some behaviors of code words. For instance, it is clear that one cannot challenge a racialized not-at-issue content through direct denial:

11) The main problem that faces the national budget is welfare.
12) No it isn't!

A speaker of 12) targets the at-issue content of 11) while failing to target the putatively racialized not-at-issue content of "welfare."

However, Stanley's theory overgeneralizes. Khoo (2017) convincingly demonstrates that the distinction between at-issue and not-at-issue content cannot account for the behavior of code words. The reason is simple: not-at-issue contents of terms and sentences do not vary from context to context. Note that speakers cannot in general cancel the not-at-issue content of an utterance without saying something infelicitous.

13) I rode my bike to the park in the rain. # It wasn't raining.
14) We ate at the restaurant Martin's mother owns. # His mother doesn't own it.

But intuitively, one can felicitously disavow commitment to the racialized connotations of "welfare" and "inner city," and it seems plausible that these terms do not even carry such racialized meanings in every context. If Stanley is correct that racialized content is not-at-issue content, 15) and 16) should sound odd, since they explicitly cancel not-at-issue contents associated with one or more terms in the discourse:

15) These tax breaks help **inner city** families. Most of them are white.
16) There are many people on **welfare**. None of them are lazy blacks.

If "inner city" and "welfare" are more or less elliptical for "lazy blacks," it should be difficult to access a coherent, felicitous reading of 15) and 16). They should even strike us as contradictory. Yet the fact that these sentences are intuitively fine strongly suggests that Stanley is wrong to locate racialized meaning in the not-at-issue content of code words/racialized terms.

The failure of Stanley's semantics is a specific instance of a more general phenomenon. Theorists usually assume that the conventional meanings illuminated by semantic theorizing are inflexible, "brittle" sorts of objects whose inflexibility is incapable of explaining the rich variation we see in the behavior of the words which express these meanings. Indeed, these are precisely the sorts of considerations that drive Khoo to his minimalistic simple theory, which is studiously non-committal about the nature of linguistic meaning. Unfortunately, we saw that that approach isn't entirely successful either: an adequate theory of racial communication in politics should mark the difference between communication by means of "true code words" and communication by means of racialized terms, and Khoo's theory does not offer those resources.

My core proposal-- that the observed racial norms at a context constrain the referential content of a racialized term-- shows that we need not choose between implausible claims about conventional meanings on the one hand or agnosticism about the nature of linguistic meaning on the other. In fact, my proposal nicely shows how we should divide the labor between linguistic theorizing and social-scientific explanation, and I will close by explaining why.

Ken Taylor (2007) has introduced the term "modificational neutrality" to describe predicates whose semantic values express properties which are neutral among particular modifications that those properties might undergo in context. For example, the predicate "red" can be modified in a number of different ways. Anything that has the property of being *red dirt* needs to be red all the way through, while anything that has the property of being a *red table* need only be red on its surface.[22] Yet importantly, we cannot and should not conclude from this that "red" means *red all the way through* when we're talking about red dirt, but means *red on the surface* when we're talking about red tables. Rather, "red" just means *red colored,* and it's a question for metaphysics (not semantics) just *in which respect* a given object needs to be red in order to fall into the extension of the predicate.[23] In effect, modificational neutrality shows that linguistic strings often express coarse-grained propositions-- Taylor's preferred term is "propositions-in-waiting"-- which then take on greater semantic determinacy in the context of interpretation.

Now, what's interesting about modificationally neutral predicates is that they are compatible with *lots* of different candidate on-the-ground modifications, and their *semantic* profile tells you very little about what modifications would count as correct at any given context. Despite the fact that semantics offers such little guidance in this department, human beings are generally pretty good at figuring out which particular modifications are most relevant at the context. For instance, humans seamlessly honor different sorts of modifications for the property expressed by "red" in a way that

[22] Equivalently: anything that properly falls into the extension of "red dirt" must be dirt that is red all the way through, while anything that falls into the extension of "red table" must be a table that is red on its surface.

[23] What we might call "radical contextualists" (e.g. Travis 1985) deny that this question is only for metaphysics, and hold that a theory of meaning must account for the various ways in which the meanings of predicates can be modified in context. Taylor's invocation of modificational neutrality is explicitly intended to show that we need not follow the radical contextualist on this point.

systematically depends on the nature of the object being described as "red"-- anything that counts as "red dirt" is red all the way through, and anything that counts as a "red chair" is red on the surface.

If we regard racialized terms as a type of community-specific speech, something like this implicit semantic knowledge is on display in the case of racialized vocabulary as well. In the right sorts of speech contexts, humans seamlessly honor different modifications for the properties expressed by "thug," "terrorist," "criminal alien," and "immigrant" in a way that depends on the racial presentation of the individuals being described. In each case, there is a *systematicity* to these modifications-- a coherent pattern between the referential content that a term is assigned in context and the racial presentation of its referents-- that calls out for an explanation.

The explanation that I offer emphasizes what is right about both Stanley and Khoo's approaches to the semantics of code words while avoiding the shortcomings of each. Stanley is right to note that code words often transcend their coded character within particular discourse communities. Yet his mistake is to think that the relative semantic stability of these not-so-coded words is explained by facts about their "conventional meaning." Khoo is right to emphasize that the notion of conventional meaning is unlikely to be helpful in understanding the high potential for variance that code words display. Yet his mistake is to think that a "simple theory" which black boxes the roles of conventional meaning and referential content in linguistic communication will be sufficient to understand relevant differences between different types of coded (and not-so-coded) speech. My approach, on the other hand, emphasizes the role of community-specific discourse norms in determining the referential content of racialized terms. This allows us to explain in one fell swoop i) why racialized terms do not achieve their cognitive effects by being "coded," ii) why the content of racialized terms seems to differ across speech contexts and discourse communities, and iii) why all this is compatible with racialized terms only bearing their "non-racialized" conventional meanings.

6. Conclusion

Let's sum up. The standard framing of racially coded speech needs an update. Whereas code words and implicit racial appeals were once practically the only game in town, the last several years have seen the rise of far less "implicit" forms of racial communication. The normative structure of mainstream racial communication has shifted in recognizable ways, and our linguistic theories should account for that shifting landscape. From a social-scientific point of view, I argued that the standard framing does not appreciate that i) there are *many* norms governing racial discourse, and that they differ in specific ways across discourse communities, and that ii) compliance with general norms of racial equality has historically waned. From a linguistic point of view, I argued that community-specific norms for the use of a term can play a role in restricting or expanding the referential content of that term at a context. We saw that this hypothesis avoids the shortcomings of approaches which accord too much or too little weight to the role of conventional meaning in accounting for the communicative function of racialized terms.

And finally, from a more programmatic point of view, I hope to have shown that understanding and modelling the re-ascendance of explicitly racist political discourse will require a number of theoretical and practical tools from a wide variety of disciplines, including political science, psychology, and theories of linguistic meaning. From this perch, perhaps we can turn to the really hard problem-- that of addressing the deeply *political* dimensions of this re-ascendance, circumspectly locating in thought and action the many cruces at which linguistic practice is bound to extra-linguistic oppression. For it is at those points that the real work will start.[24]

Oakton Community College
podonnell@oakton.edu

References

Blauner, Bob. "Talking past each other: Black and white languages of race." *The American Prospect* 10 (1992): 55-64.

Blow, Charles. "Paul Ryan, Culture, and Poverty." *The New York Times*. March 21, 2014..

Bonilla-Silva, Eduardo. *Racism without racists: Color-blind racism and the persistence of racial inequality in America*. Rowman & Littlefield, 2017.

Craig, Maureen A., and Jennifer A. Richeson. "More diverse yet less tolerant? How the increasingly diverse racial landscape affects white Americans' racial attitudes." *Personality and Social Psychology Bulletin* 40.6 (2014a): 750-761.

Craig, Maureen A., and Jennifer A. Richeson. "On the precipice of a "majority-minority" America: Perceived status threat from the racial demographic shift affects White Americans' political ideology." *Psychological Science* 25.6 (2014b): 1189-1197.

Diaz Legaspe, J., 2018. Normalizing Slurs and Out-group Slurs: The Case of Referential Restriction. *Analytic Philosophy*, *59*(2), pp.234-255.

Dice, Mark. "Trump speech to police officers- law and order will return to streets of America." *Video*. Uploaded to YouTube July 28, 2017.

[24] Thanks to organizers and audiences at the 2018 Princeton Workshop in Social Philosophy, the 2018 Language and Power workshop in Barcelona, and the 2019 APA-Eastern Divisional meeting in New York. Special thanks to Sam Berstler and Rachel McKinney, who each gave extensive and constructive comments on earlier drafts. Their contributions have made this paper much better than it would have been otherwise. Thanks also to Steven Gross, David Lindemann, and Daniel Estrada for helpful discussion, and to several anonymous reviewers for comments, questions, and criticisms that greatly improved the paper. Finally, a conversation with Ken Taylor a few months before his untimely passing was largely responsible for the development of the positive view articulated here, and so thanks are due to him. While it is undoubtedly a slight consideration compared to those other losses that his death has meant, I regret that I will not be able to revisit these topics with Ken.

Hebert, Cassie and Rebecca Kukla. "Ingrouping, Outgrouping, and the Pragmatics of Peripheral Speech." *Journal of the American Philosophical Association* Vol. 2, Issue 4 Winter 2016 pp. 576-596

Huber, Gregory A., and John S. Lapinski. "The "race card" revisited: Assessing racial priming in policy contests." *American Journal of Political Science* 50.2 (2006): 421-440.

Huber, Gregory A., and John S. Lapinski. "Testing the implicit-explicit model of racialized political communication." *Perspectives on Politics* 6.1 (2008): 125-134.

Hurwitz, Jon, and Mark Peffley. "Playing the race card in the post–Willie Horton era: The impact of racialized code words on support for punitive crime policy." *Public Opinion Quarterly* 69.1 (2005): 99-112.

Hutchings, Vincent L., Hanes Walton Jr, and Andrea Benjamin. "The impact of explicit racial cues on gender differences in support for confederate symbols and partisanship." *The Journal of Politics* 72.4 (2010): 1175-1188.

Kertscher, Tom. "Were Paul Ryan's poverty comments a 'thinly veiled racial attack?" politifact.com. March 14, 2014.

Khoo, Justin. "Code words in political discourse." *Philosophical Topics* 45.2 (2017): 33-64.

Kinder, Donald R., and Lynn M. Sanders. *Divided by color: Racial politics and democratic ideals*. University of Chicago Press, 1996.

Lind, Dara. "Trump on deported immigrants: 'They're not people. They're animals.'" *Vox*. May 17, 2018

Leader Maynard, J. and Benesch, S., 2016. Dangerous speech and dangerous ideology: An integrated model for monitoring and prevention. *Genocide studies and prevention: an International Journal*, *9*(3), p.8.

Marques, Teresa. "" Beasts in human form": How dangerous speech harms." *Araucaria* 21, no. 42 (2019).

Mendelberg, Tali. *The race card: Campaign strategy, implicit messages, and the norm of equality*. Princeton University Press, 2001.

Mendelberg, Tali. "Racial priming revived." *Perspectives on Politics* 6.1 (2008): 109-123.

Morrison, T., 1993. On the backs of blacks. *Time*, *142*(21), pp.57-57.

Mühlebach, D., 2019. Semantic contestations and the meaning of politically significant terms. *Inquiry*, pp.1-30.

Perlstein, Rick. "Exclusive: Lee Atwater's Infamous 1981 Interview on the Southern Strategy." *The Nation*, Nov. 13, 2012.

Saletan, William. "What Trump supporters really believe." Slate.com. August 29, 2017.

Saul, Jennifer M. 2015. "I am an immigrant-- and I believe 'migrant' is a far from neutral term." *New Statesman.*

Saul, Jennifer M. "Racial Figleaves, the Shifting Boundaries of the Permissible, and the Rise of Donald Trump." *Philosophical Topics* 45.2 (2017): 91-116.

Schwartz, Ian. October 10, 2017. "Bannon declares war on GOP establishment: 'We're coming after all of them and we're going to win." *RealClearPolitics.*

Smith, D.L., 2011. *Less than human: Why we demean, enslave, and exterminate others*. St. Martin's Press.

Stanley, Jason. *How propaganda works*. Princeton University Press, 2015.

Taylor, Jessica. "Trump tests midterm message on immigration, MS-13 'animals' during Tenn. rally" *NPR* May 29, 2018.

Taylor, K., 2007. A little sensitivity goes a long way. *Context-sensitivity and semantic minimalism: New essays on semantics and pragmatics*, pp.63-93.

Tirrell, L., 2012. Genocidal language games. *Speech and harm: Controversies over free speech*, pp.174-221.

Travis, C., 1985. On what is strictly speaking true. *Canadian Journal of Philosophy*, *15*(2), pp.187-229.

White, Ismail K. "When race matters and when it doesn't: Racial group differences in response to racial cues." *American Political Science Review* 101.2 (2007): 339-354.

Wodak, Ruth. *The politics of fear: What right-wing populist discourses mean*. Sage, 2015.

Wodak, Ruth. "The 'Establishment,' the 'Elites,' and the 'People': Who's Who" J*ournal of Language and Politics*, Vol. 16, No. 4, 2017, p. 551-565.

Yin, T., 2013. Were Timothy McVeigh and the Unabomber the Only White Terrorists: Race, Religion, and the Perception of Terrorism. *Ala. CR & CLL Rev.*, *4*, p.33.