Linguistic Hijacking

Derek Anderson
Boston University
deanders@bu.edu

Linguistic Hijacking[1]
Derek Anderson

Abstract

This paper introduces the concept of linguistic hijacking, the phenomenon wherein politically significant terminology is co-opted by dominant groups in ways that further their dominance over marginalized groups. Here I focus on hijackings of the words "racist" and "racism." The model of linguistic hijacking developed here, called the semantic corruption model, is inspired by Burge's social externalism, in which deference plays a key role in determining the semantic properties of expressions. The model describes networks of deference relations, which support competing meanings of, for example, "racist," and postulates the existence of deference magnets that influence those networks over time. Linguistic hijacking functions to shift the semantic properties of crucial political terminology by causing changes in deference networks, spreading semantics that serve the interests of dominant groups, and weakening the influence of resistant deference networks. I consider an objection alleging the semantic corruption model gets the semantic data wrong because it entails those who hijack terms like "racist" speak truly, whereas it's natural to see such hijacking misuses as false speech about racism. I then respond to this objection by invoking the framework of metalinguistic negotiation proposed by Plunkett and Sundell.

---

1. What Is Linguistic Hijacking?

This paper identifies a form of epistemic and political violence I call *linguistic hijacking*. Linguistic hijacking involves misusing or co-opting politically significant terminology in ways that harm marginalized groups. These misuses are often intended to exert a kind of political dominance, though they do not need to be intentionally controversial or politically motivated to count as hijackings. The central aim of this paper is to investigate linguistic hijacking by theorizing about the harms it causes. The present study engages with a complex set of issues in epistemology, critical race theory, and metasemantics.[2]

Linguistic hijacking occurs when dominant groups misuse terminology that marginalized groups use to communicate and theorize about the forms of oppression they experience. Something counts as a *mis*use of such politically significant terminology insofar as it functions to spread misinformation, ignorance, or false belief about the phenomenon referred to within the relevant marginalized discourse. For example, misuses of the word "racism" are those that function to spread misinformation, ignorance, or false belief about racism.

Linguistic hijacking is more than a mere misapplication of a term; it is a pattern of inaccurate usage that functions as part of a resistance to knowledge of oppression. Resistance to knowledge of oppression is a form of active ignorance (Collins 2002; Mills 2007; Pohlhaus 2012; Sullivan and Tuana 2007), a multifaceted group activity aimed at erasing or containing the spread of knowledge about things like racism, patriarchy, transphobia, homophobia, capitalist exploitation, ableism,

---

[2] Metasemantic theories describe the conditions that determine the semantic features of a language (fragment). Where a semantic theory describes the semantic values of expressions and the rules by which the semantic values of complex expressions are determined by the features of their parts, a metasemantic theory explains why the correct semantic theory is true, rather than some other semantic theory. A metasemantic theory might also provide counterfactual descriptions of what the correct semantic theory would have been if other facts—psychological, social, historical, physical, perhaps others—had been different. This distinction has been drawn many times; see Hawthorne (2007), Williams (2007), Williamson (2008), Rayo (2010), Sider (2013), Yalcin (2014), and Simchen (2017). The use of the term "metasemantics" to mark the foundational side of this distinction is only recently established. Earlier authors draw roughly the same distinction using different terminology. Stalnaker (1997) distinguishes between "descriptive semantics" and "foundational semantics." Schiffer (1987) talks about "semantic theories" and "meaning theories" (for Schiffer, meaning theories include semantic theories as components). For Lewis (1969) there is the study of "languages" and of "what makes a language be an actual language of a population."

and so on, in order to protect and perpetuate systems of domination. It involves spreading misinformation, misrepresenting crucial facts and concepts, misdirecting attention, covering up evidence, erasing testimony and silencing voices of oppressed groups, creating counternarratives, and many other things as well. A particular use of a term, such as a misapplication of "racism," counts as linguistic hijacking if it is part of a usage pattern that facilitates resistance to knowledge of oppression in this sense.

Understanding linguistic hijacking is important. It is a matter of understanding the ways in which misuses of politically significant terminology function to undermine the effectiveness of the conceptual resources deployed by marginalized groups in political discourse. On the model developed in this paper, linguistic hijacking allows dominant communities to isolate themselves from marginalized discourses and ignore the realities of systemic oppression. Further, I show how pervasive misuse might actually destabilize the semantic functions of crucial terminology within marginalized discourses and within broader public conversations. For example, widespread misuse of the word "racism" might ultimately corrupt the semantic function of that word, shifting its reference to something other than racism. If the word "racism" ceases to refer to racism, then an invaluable tool in the fight against oppression has been destroyed.

Linguistic hijacking is perhaps best introduced through examples. The primary cases I examine here are hijackings of the words "racism" and "racist." Consider the white protestor holding a sign that says, "Affirmative action is racist," or the angry Twitter account that tweets "#BLM is a racist organization." The first implies that a practice that disadvantages white people on account of their race is a form of racism against whites. The second implies that an organization concerned to protect the lives of Black people, but not explicitly concerned to protect the lives of white people, is racist. In each of these examples, the word "racist" is deployed in a way that obscures the nature of racism, invoking an alternative conception on which any differential treatment aimed at promoting the well-being of members of an explicitly specified race counts as "racist."

Affirmative action explicitly aims to promote the chances of people of color in relation to their white competitors, but to call this "racist" obscures the nature of racism. It misdirects moral opposition aimed at racist institutions and turns that opposition against institutions designed to combat racism, while reinforcing resistance to knowledge of racism. A closely related misuse of "racist" alleges that conversations about the detail and extent of white supremacy are "racist." According to this misconception, sometimes called 'colorblindness,' any reference to race is a form of "racism," so any reference to white people or whiteness as such counts as "racist." Hence, any attempt to explain a racist incident by referencing

whiteness is seen as "racist."[3] For example, Richard Dawkins once tweeted that it was "racist" to call him a white man.[4]

Such statements use the words "racism" and "racist" in ways that do not accord with our best theories of racism. Racism is a matter of power and privilege, not merely a matter of difference in treatment or concern on the basis of race. Collins (1998, 280) theorizes racism as:

> A system of unequal power and privilege in which human beings are divided into groups or races, with social rewards being unevenly distributed to groups based on their racial classification. Variations of racism include institutionalized racism, scientific racism, and everyday racism. In the United States, racial segregation constitutes a fundamental organizing principle of racism.

Collins's definition reflects extensive theoretical work within Black feminism (Lorde 2012; hooks 1981; Davis 2011; Collins 1986), critical race theory (Lawrence 1987; Crenshaw 1988; Peller 1990; Gotanda 1991) and intersectionality studies (Crenshaw 1991, 2018; Roberts 1991; Cho, Crenshaw, and McCall 2013). Crucial for present purposes, racism is a matter of unequal power and privilege maintained by a *system*, something of larger ontological scale than an individual.

Affirmative action aimed at promoting the success of people of color is not racist on this view of racism; nor is Black Lives Matter a racist institution in virtue of its explicit mission to promote the life chances of Black folks; nor is it racist to identify whiteness as implicated in patterns of racism. I always use the word "racism" to refer to racism, so understood. I use the word "racist" to express the property of being racist, which I take to be the property of promoting or beneficially participating in racism.

---

[3] The idea that differential treatment on the basis of race in the service of addressing past racial injustices should itself count as racism has been subjected to thorough criticism by scholars of critical race theory. See Kennedy (1990) and Gotanda (1991) for canonical arguments.

[4] Dawkins tweeted, "'insufferable smug white male making snide comments in loafers.' Racism & sexism are fine, so long as they point in the right direction!" (@RichardDawkins, Twitter, May 23, 2013, 11:40 p.m. https://twitter.com /RichardDawkins/status/337775059139694593), and "Learn to think clearly and use language precisely. You may JUSTIFY racism & sexism towards white males. But it's still racist & sexist." (May 24, 2013, 6:04 a.m. https://twitter.com/RichardDawkins /status/337871610826600448).

This discussion of racism highlights an important feature of theorizing about linguistic hijacking: to classify a particular usage of an expression as an instance of linguistic hijacking presupposes something about the underlying metaphysical reality of the first-order phenomena in question. It presupposes the veracity of some marginalized discourse. In the present study, a theory of racism is presupposed—namely, the theory that racism is essentially related to structural oppression. Not everyone accepts such a theory of racism, and this difference tends to track political convictions related to race. Consequently, claims about linguistic hijackings of "racism" will be politically contentious along roughly the same lines that the underlying theory of racism is contentious. That is, a person who disagrees with me about the first-order facts about racism will also very likely disagree with me about which usages of the word "racism" count as cases of linguistic hijacking. This is a general fact about linguistic hijacking. Whether some usage counts as linguistic hijacking is theory-dependent. It always depends on background social theories of oppression and privilege.

The fact that there are predictable disagreements about racism—predictable because they reliably track differences in political commitments—does not entail there is no objective fact about what racism really is. Likewise, there are politically situated disagreements about which uses of "racism" count as linguistic hijackings, but this does not entail there is no objective fact of the matter about whether a given use counts as a linguistic hijacking. There are misuses of "racist" that function to spread misinformation and facilitate white ignorance (Mills 2007) about racism; I maintain that this is objectively true, even though it is disputed across familiar political divides. Some white supremacists may claim that white people are more racially oppressed than people of color in the United States these days. These same white supremacists may claim that "racism" is being hijacked by critical race theorists. But the white supremacists are wrong on both counts. Assessing whether a given usage counts as linguistic hijacking requires applying a background theory of oppression, but this does not mean the concept of linguistic hijacking is subjective or somehow problematically biased.

Other examples of linguistic hijacking include attacks on gender identity, such as the statement "Trans women are not really women." Here the term "women" is used in an attempt to exclude trans women from its extension. The correlated misuse of "men" functions in the same way, as when Milo Yiannopoulos said, "I think that women and girls should be protected from having *men* who are confused about their sexual identities in their bathrooms."[5]

---

[5] *Real Time with Bill Maher*, HBO, episode 415, February 17, 2017; emphasis added. For extensive discussion of the semantics and politics of uses and misuses of gender terminology, see Bettcher (2009, 2013).

Misuses of political organizing terms such as "feminism" and "intersectionality" are also important examples of linguistic hijacking. Examples of mischaracterizations of feminism within mainstream political discourse are commonplace and widely recognized by feminist scholars and activists. One can get a sense of these by searching "feminism" on the *Urban Dictionary* website, where one will find such definitions as "Feminism: A relentless political advocacy group pushing for special privileges for women, which pretends to be a social movement advocating equal rights for the genders,"[6] and "Feminism: An ideology that supports that women are superior to men, whether biologically, socially, or mentally."[7] These definitions flagrantly misconstrue and misrepresent what feminism is and help to spread misinformation.

Hijackings of "intersectionality" tend to be situated in white academic discourses. Henning (2020) gives an example in which a scholar attempts to quantify intersectionality using a statistical model, but the model adduced includes properties such as height and t-shirt preference in addition to race and gender. As Henning argues, intersectionality is not concerned with 'identities' such as t-shirt preference, which have no connection to systemic oppression. The attempt to deploy the word "intersectionality" for a model of such properties functions to obscure knowledge of intersectionality properly understood, which is essentially centered on interlocking systems of oppression. According to this perspective, any attempt to use the term "intersectionality" for identities that are not centrally connected with systemic oppression is a form of linguistic hijacking.[8]

There are many important cases of linguistic hijacking, each of which presents a unique constellation of issues. While my focus will be on hijackings of "racist" and "racism," my conclusions bear on epistemic and metasemantic issues concerning a very wide range of expressions involved in political speech.

Linguistic hijacking occurs when dominant epistemic agents appropriate an expression that is epistemically relevant to the pursuit of social justice in a way that furthers the oppression of some marginalized group. Take the term "racist." This term is epistemically relevant to the pursuit of justice insofar as it is our primary linguistic representation of racism. We use it to communicate and theorize about

---

[6] *Urban Dictionary*, s.v. "Feminism," accessed July 30, 2020, https://www.urbandictionary.com/define.php?term=feminism.

[7] *Urban Dictionary*, s.v. "Feminism," Definition #12 on Sept. 7, 2019, https://web.archive.org/web/20170909091321/http:/www.urbandictionary.com/define.php?term=Feminism&page=2.

[8] For further discussion of what Collins (2015) calls "the gentrification of intersectionality," see Tomlinson (2013), Cho, Crenshaw, and McCall (2013), and Collins and Bilge (2016).

racism, to direct social and political action in opposition to racism, to challenge oppressive people and institutions, and so on. When "racist" is appropriated by political opponents of antiracist theory and activism, it's redeployed in ways that obscure or twist the meaning of the word. It's used to attack antiracism advocacy groups such as Black Lives Matter, labeling them as "racist." It's used in ways that obscure the nature of racism by applying the term to that which is not racist.

Insofar as linguistic hijacking systematically infringes on the epistemic agency of marginalized groups, it is a form of epistemic oppression (Dotson 2014). It might also be conceived as a form of epistemic injustice (Fricker 2007; McKinnon 2016) insofar as it can be connected with individual or group-based wrongs against members of marginalized groups in their capacity as agents of knowledge.[9] Linguistic hijacking can facilitate willful hermeneutical ignorance (Pohlhaus 2012), a form of active ignorance through which dominant agents resist uptake of conceptual resources generated by marginalized groups; this is a crucial way in which linguistic hijacking functions to promote active ignorance, as will be discussed below. Linguistic hijackings of "racist" and "racism" also contribute to what Mills (2007) calls *white ignorance*, the massive, widespread miscognition among white people about the nature and extent of racism. As will also be discussed later in more detail, linguistic hijacking helps white folks construct misleading discourses about what they call "racism," but which is not really racism. This ultimately allows white folks to avoid thinking or talking about racism.

How does linguistic hijacking work? In section 2, I introduce what I call the *semantic corruption model*. On this model, linguistic hijackings alter the semantic properties of politically significant terminology, changing them so they no longer function semantically in ways that promote knowledge and communication about oppression. This dynamic metasemantic model gives us a powerful tool for understanding how patterns of abuses of terminology might affect meanings in epistemically harmful ways. Section 3 then presents an objection to the model—the semantic data objection—which focuses on how the model renders dominant agents as speaking truly in their own idiolects about something other than racism, despite the fact that it is quite natural (from the perspective of the critical race theorist, at least) to represent them as speaking falsely about racism. Section 4 responds to this objection, drawing on the idea of metalinguistic negotiation developed by Plunkett and Sundell (2013). Section 5 concludes with a brief discussion of the social nature of metalinguistic negotiation as it operates according to the semantic corruption model.

---

[9] Although see Dotson (2012) on the limits of epistemic injustice as a framework for understanding epistemic oppression.

2. The Semantic Corruption Model

Semantic corruption is a manipulated semantic change imposed by those in power that furthers their ends, quintessentially by debasing important connections between representation and reality. By misusing the words "racist" and "racism," the dominant group seizes control of the meanings of those terms, altering them to serve its own interests. This allows dominant groups to shelter themselves from perspectives that would challenge the status quo. The quintessential example of this is when white people use "racist" to refer to initiatives that impose burdens on white people in the course of redressing racist inequality. Linguistic hijacking also creates metasemantic echo chambers. This happens when communities of white folks reaffirm one another's mistaken usage of "racism" and refuse to uptake the meaning deployed by those who experience racism. And it has the potential to disrupt thought and communication about racism by eroding the semantic properties of expressions that are used within marginalized discourses, threatening to deprive them of their semantic function within those discourses and within broader conversations. This happens when linguistic hijacking of the word "racism" causes a change of meaning so that it no longer refers to racism, possibly even among those who are interested in talking about racism.

A natural way to model the idea of semantic corruption is to draw on Burge's (1979, 1986) theory of social externalism, according to which patterns of deference determine the semantic properties of mental and linguistic representations.[10] As persuasive members of dominant groups misuse the term "racism" and as this usage becomes more popular—as more people defer to those who engage in it—the term's meaning gradually aligns with dominant usage. Eventually we can no longer use "racism" to talk about racism. Its original meaning has been corrupted. The term is now co-opted by the oppressor. It expresses a concept of the oppressor's choosing.

---

[10] The present paper does not attempt a thorough defense of Burge's social-externalist framework. There are many complex issues regarding the nature of semantic deference and the adequacy of a semantic-externalist framework of the kind developed here, as well as potential disputes about how Burge himself intended these issues to be understood and resolved. Certainly, many papers could be written to explore such foundational issues. My aim in the present paper is to analyze a less foundational phenomenon, viz., linguistic hijacking, using Burge's externalist insights as a starting point.

2.1. A Burgean Thought Experiment

Let's start by reimagining the classic Burgean thought experiment supporting social externalism. Rather than focusing on arthritis as Burge (1979) does, we will focus on racism.

Imagine Eric, a young white man who thinks racism is only a matter of treating people differently on the basis of race. Eric explicitly maintains that racism against white folks is possible and happens fairly often. Eric frequently and whole-heartedly asserts, "Affirmative action is racist against white people." Given that racism necessarily involves systematic oppression white people don't face, Eric's claim is false. Eric is also expressing a belief about racism that is likewise false.

Next imagine Eric has never come across anyone who disagrees with him on this point until he attends a university where he eventually pursues a degree in sociology with a focus on critical race theory and intersectionality studies. On his first day of Sociology 101, he comes right out with his claim that affirmative action is racist. The sociology professor rebuts him, explaining in some detail why affirmative action cannot be a form of racism because affirmative action does not reinforce systems of racial oppression. By the time he finishes his degree, Eric takes himself to have had a change in belief about racism spurred by conversations with his professors and classmates and by reading critical race theory. He stands corrected. He concedes that his earlier belief was false. He defers to his professors and classmates in their usage of the term "racism." It's quite natural to interpret this as a case in which his original belief was false, as was the sentence he used to express that belief.

Now consider a different possible world in which patterns of systemic oppression are exactly as they are in our actual society—leaving facts about racism unchanged—but in which the term "racism" is not used to refer to racism but is used instead to refer to the property of merely discriminating on the basis of race. In this world when Eric says affirmative action is "racist," his Sociology 101 professor agrees. All of the critical race theory he reads agrees with him on this point. Eric takes himself to have been correct all along in his use of that term.

Following Burge's intuition about arthritis, we conclude that here Eric's sentence "Affirmative action is racist against white people" and the belief it expresses are both true given the way people use the word "racist" in this alternate possible world. Since none of the facts about racism are different in this other world, we conclude that the difference in truth-value must consist in a semantic difference. The semantic features of the word "racist" and the concept expressed by that term are different in this other world; relatedly, the belief Eric expresses is not about racism.

Now let's modify the Burgean thought experiment so there are a number of different communities in the same world, some who use the word "racism" for

racism and some who use it for mere race-based discrimination. We imagine that both "racist" and "racism" have correspondingly different meanings within each community.[11]

This scenario prompts the following thought. What Eric means by "racism" and the concept he thinks about in connection with that term are determined by which group he is disposed to defer to. If upon reflection Eric defers to the critical race theorists, then he thinks and talks about racism. If he only defers to speakers who maintain that "racism" in twenty-first-century America is primarily directed against whites, then Eric does not think or talk about racism. He only thinks and talks about some other phenomenon that is not essentially connected with racial oppression. When Eric believes and says, "Affirmative action is racist against white people," he speaks and believes truly, though his speech and thought are not about racism.

In order to model this relationship between meaning and deference, we have to think carefully about the ways in which members of a community are related to one another in terms of who defers to whom. Each person has a unique deference relation with each other person in the community. These dispositions to defer also vary on an expression-by-expression basis. I may be willing to defer to a

---

[11] It is stipulated as part of the Burgean thought experiment that the counterfactual community uses the word "arthritis" to express a different concept with a different extension, such that the truth conditions of "I have arthritis in my thigh" are different in the counterfactual world. Burge doesn't explain what the metaphysical ground of this difference is, and I follow him in that. Likewise, in my thought experiment, it's stipulated that the counterfactual community uses "racism" to express a different concept with a different extension. In the second version of the thought experiment, it's stipulated that this alternative community exists in the actual world alongside the other; hence it's stipulated that the two uses of "racism" express different concepts and have different extensions.

This highlights the questions: where did these meanings come from in the first place? How came there to be two groups using one word with two different meanings? I leave these details unspecified here. There are many different kinds of accounts that could be given. The two communities might associate the term "racism" with different conceptual roles (Greenberg and Harman 2008), or they could be so constituted that their concepts have different asymmetric dependence properties (Fodor 1990), or the concepts they use carry different information (Dretske 1991) or differ in "direct proper function" (Millikan 1984). Burge (1979) expresses sympathy with an antireductionist view of semantics; on such a view, any explanation of why the two communities mean different things must make reference to what members of each community *mean* when they say "racism."

mathematician about the meaning of "abelian group" but not "axolotl." Hence, for every expression E in the language, each person S has their own unique deference disposition profile—a set of relationships to each other person S' in the community characterized by S's willingness to defer to S' concerning the correct use of E.[12]

Note that (following Burge) deference is grounded in dispositions rather than interactions. S is deferential to S' just in case S is *disposed* to be convinced by S' of the correct usage of E. This may be true even if the two never actually interact. We can say the strength of the deference relation from S to S' concerning E is given by P in the following counterfactual: if S were to interact with S' about the usage of E, the probability that S would defer to S' concerning E is equal to P. Hence a measure of deferential strength is a measure of a certain probability within a counterfactual situation.[13]

2.2. Constructing the Model

We are now primed to give a more exact specification of the semantic corruption model. We represent each individual in the broader language-using community by a node in a network. An individual's deference profile for a given term E is represented by a set of weighted arrows pointing to every other node in the network. These arrows indicate the strength of deference the individual bears to their fellow regarding E. A complete specification of a person's deference profile at a given time would include a set of indexed layers of arrow maps, one layer for every term. This allows the model to discriminate deferential attitudes concerning every term the individual recognizes. We may identify the index of each layer with the term it represents; so for example, "racism" would be the index of its own layer. The total deference pattern for a single term across members of a community is given by the full collection of arrow maps for the relevant index, where each node in the network is associated with its own co-indexed arrow map. The set of all co-indexed arrow maps captures the full range of influence each node has on every other concerning the proper use of a single term at a time. It provides a static

---

[12] The strength of these deference relations may be indeterminate in some cases, vague in many or perhaps all cases, and also context sensitive. The model I develop here idealizes away from these nuances, which are not crucial for understanding the semantic corruption model but which may be of significance when applying the dynamic metasemantic framework in detail.

[13] If the suppression of context sensitivity seems too acute at this point, we can define the measure of deference as an average of the probabilities across sufficiently close possible worlds in which an interaction between S and S' concerning E takes place. A more detailed account is beyond the scope of the present paper.

representation of, for example, who is disposed to defer to whom concerning the semantics of "racism" at a single moment.

Let's suppose there are two potential semantic interpretations for "racism" available, one referring to racism-as-oppression and one referring to race-conscious differential treatment. We represent the degree to which each person is committed to each interpretation by assigning two scalars to each node within the "racism" layer. We might graphically represent these numbers as colors for each node with one scalar fixing a level of red saturation and the other a level of blue saturation. Fully blue nodes stand for people who use "racism" to refer to racism. Fully red nodes stand for people who use "racism" to refer to race-conscious differential treatment with or without systemic oppression. Purple nodes are in semantic limbo, being disposed to defer more or less evenly to members of the two conflicting semantic camps.[14] The scalars at node N are determined as a function of the scalars at each other node N_i together with the weight of N's deference arrow to N_i. A node that defers strongly to people with high blue saturation and defers weakly to nodes with high red saturation will be on the blue end of the spectrum—that is, they will use "racism" to denote racism—while a node that defers strongly to red nodes and weakly to blue nodes will be red and use "racism" to denote mere differential treatment on the basis of race.

People with highly determinate deference profiles about "racism"—those whose nodes are not purple—will form subnetworks, collections of nodes that point strongly at one another and point only weakly or with degree zero to nodes of the opposing color. These subnetworks represent communities that are strongly disposed to defer to one another concerning the semantics of "racism" and that are unwilling to defer to members of opposing semantic factions. Thus, this model produces a natural demarcation of linguistic subcommunities who use terms in accord with one another and who collectively oppose the usage patterns of outsiders. The model also grounds the claim that members of different subcommunities mean different things by "racism."

The semantic corruption model entails a kind of social constructivism about truth. Each linguistic subcommunity has the power to determine for itself the semantic properties of its linguistic representations by determining who should and who should not be deferred to. By deferring only to those who use terms in preferred ways—that is, by deferring only to one another—a community has the

---

[14] What does it mean to be in semantic limbo? A purple node in the "racism" layer has no clear conviction about what racism is or what "racism" means, such that they are disposed to defer to critical race theorists as well as their opponents in equal measure. Perhaps such a person does not determinately express any meaning when they use "racism." Thanks to Zeynep Soysal for raising this question.

power to stipulate the truth of whatever statements it cares to. A reactionary community can thereby stipulate that all of its proclamations about "racism" are true. The ontological features of reality are not altered—the community does not make racism itself vanish. Only the semantic features of their language are manipulated.

The possibility of socially constructing truth extends to mental representations as well, in a sense. Once reactionary communities convince each other that "Affirmative action is racist" is true, the belief expressed by this sentence within that community is also true. The belief would not be about racism. In this way, the reactionary community leads itself to stop thinking falsely about racism and believe only truths about a gerrymandered set of facts that excludes facts about racial oppression, white privilege, or any other aspects of true racism. The community refuses to think about those uncomfortable facts. Thus, dominant groups are free to dissociate from uncomfortable swaths of reality by constructing their own meanings for social and political discourses.

We can now understand the harm of hijacking "racism" according to this dynamic metasemantic theory (note, we have not yet explicitly introduced the dynamic component; that will be introduced explicitly in section 2.3). Hijackings undermine deferential support for those who use the term "racism" to refer to racism and increases deferential support for those who use "racism" to refer to something else. When someone convincingly asserts, "Affirmative action is racist against white people," they strengthen deferential relations within the network that refuses to talk about actual racism. They also put pressure on deferential networks of people of color and their allies who use "racism" to think and talk about racism. Hijacking thus threatens to undermine any usefulness the term "racism" has for promoting social justice. It might even eliminate the concept of racism as a cognitive resource altogether by cutting off all means of expressing and communicating that concept. If the word "racism" ceases to refer to racism, then an invaluable tool in the fight against oppression has been destroyed.

On some versions of the semantic corruption model it may be possible for marginalized groups to persist in using "racism" to refer to racism indefinitely by refusing to defer to dominant misuses. They would maintain their own counterhegemonic subnetworks. In that case, people of color and their white allies would be able to think and talk about racism amongst themselves but not as part of the larger community in which they are embedded. While this semantic defeat would be less dramatic than the complete obliteration of "racism" as a word for racism, losing power to think and communicate about racism in shared public discourses would still be quite devastating. This would be a scenario in which, for the most part, even antiracism advocates and theorists have abandoned using "racism" to refer to racism, and though there are some small groups who continue

to use it thus, their usage is so at odds with the mainstream discourse that they are seen to be confused or wrong by definition, even among political allies. All previous discourses around racism centering on a critique of power and oppression thereby come to be seen as confused. Whatever new term is introduced for racism (assuming there is one), it will not have a historical connection to antiracist movements of the sort that using consistent terminology provides, nor will new theories constructed using that term be seen as elaborations of earlier theories of racism (insofar as those earlier theories will be seen as having a different subject matter). Lastly, if the term "racist" is co-opted by the forces of white ignorance, they can use that word to claim that opposition to white supremacy is "racist" whenever it confronts whiteness or white people as such over their role in continuing race-based oppression.

2.3. Modeling Change over Time

The set of all co-indexed arrow maps for all terms represents one static moment of life in the community: the total pattern of deference at a time *t*. The semantic corruption model represents more than this one moment. It represents a dynamic system in which meanings change over time. The dynamic model describes how hijackings alter the distribution of deference-arrow magnitudes or, more generally, how use influences patterns of deference and thereby influences patterns of semantic features.

Those who are perceived as having the greatest competence elicit the most deference. Burge describes *greatest competence* as consisting "in abilities to draw distinctions, to produce precisifications, to use numerous linguistic resources, to offer counterexamples to proposed equivalences—that elicit the reflective agreement of other competent speakers. . . . To put it crudely, a person counts as the *most competent* if he or she would be *persuasive* to other competent speakers in the use and explication of the language" (Burge 1986, 702; emphasis in original). As members of each subcommunity attempt to persuade people, those perceived as most competent will act as *deference magnets*, drawing more and more nodes into their cluster and perhaps weakening the deferential resolve of tenuous members of opposing factions.[15]

To represent the effects of deference magnets, a dynamic model might be constructed as follows. Treat each static graph as a state of a phase space containing

---

[15] This analysis positions the semantic corruption model as a version of what Ball (2020, 5) calls "power metasemantics," on which "the experts are simply those whose testimony we will accept—those who can convince us to adopt their usage," although the semantic corruption model complicates the use of the term "we" here, since different groups will recognize different "experts."

all possible patterns of deference. Within each state, the dynamic model assigns a matrix to each node representing the force with which that node acts as a deference magnet on all other nodes. Each matrix represents the actions, speech acts, prestige, credibility, and so on, of the person represented by the node—whatever features determine that person's perceived competence for others. The dynamic model then represents the effects of this perceived competence on deference patterns as a propagation operator that determines the next phase of the system as a (possibly nondeterministic) function of the effects of deference magnets together with the deference profiles of all nodes in the current phase.

We could also build in as a feature of this model that as a particular usage achieves greater coverage of the graph, the pressure on nodes to conform to the dominant usage increases. This would be represented by a global variable that functions to increase the strength of deference magnets who use a term in accord with popular usage and decrease the strength of deference magnets who use it in accord with some unpopular usage. This would represent the difficulty of getting people to use a word in a way that runs counter to popular usage. It would also reflect the idea that very idiosyncratic interpretations of terms will tend to be extinguished out of sheer impracticality. In the scenario imagined above, where a small group continues to use "racism" to refer to racism in spite of widespread rejection of that meaning, the group successfully resists the pressure represented by this global variable. Doing so requires strong resolve, since almost no one defers to you, and the vast majority, even many of your friends and allies, insist you are misusing the word.

2.4. Harms of Linguistic Hijacking

This subsection outlines a number of ways in which linguistic hijacking, understood according to the semantic corruption model, creates active ignorance and misinformation and generates other harms. It also discusses ways in which resistant groups use deference networks to generate and distribute resistant meanings. I begin with a consideration about why hijacking "racism" is especially harmful insofar as it turns a resource for alleviating oppression against the very people it is meant to help.

The negative moral valence of the word "racism" derives from the fact that it refers (and has referred) to racism. It is because "racist" denotes racists that it's bad to be called "racist." This matters to the force of the complaint that affirmative action is "racist," that it's "racist" to call Richard Dawkins white, and so on. When you condemn a person or institution using the word "racist," you are morally condemning with force that is proportional to the badness of racism. You are lumping your target in with Nazis and slave owners and the KKK. When a white person calls affirmative action "racist," they are attempting to invoke the moral

outrage proper for racism, co-opting that outrage and using it in defense of the system that undergirds white dominance. This is a crucial aspect of the manipulation of the semantics of "racism," for the change of meaning brought about through semantic corruption does not alter the moral valence associated with the term (at least not quickly). This is why it is strategically useful for white people to co-opt the term for their own purposes. When a white person says a person of color is "racist" against white people, they draw on the moral condemnation appropriate for racism and turn it against someone who actually experiences and perseveres against racism.

If the semantic corruption model is correct, we can expect what we might call *enfranchised semantic drift*, the phenomenon in which the dynamics of meaning-change will tend to favor the interests of dominant groups. Enfranchised semantic drift is likely, on the semantic corruption model, due to pervasive patterns of conceptual competence injustice (Anderson 2017). Conceptual competence injustice occurs when a member of a marginalized group is judged to have less conceptual competence than she in fact has. Such a person is unjustly regarded as lacking an adequate understanding of how a word is properly used or as failing to grasp the concept that word expresses. Because members of marginalized groups are accorded less intellectual credibility, they are consistently exposed to conceptual competence injustice.

Since deference magnetism is a matter of perceived competence, systematic conceptual competence injustice reliably reduces the power of marginalized deference magnets. Hence, patterns of deference toward marginalized communities will be relatively weak compared to patterns of deference toward enfranchised communities. For example, pervasive dominant ideology ensures that white people will typically command more respect and wield more intellectual authority than people of color. Thus, according to the semantic corruption model, if left alone over time, the system will tend to settle the semantics of the term "racism" in favor of the preferred usage of white people. Since white people also have a vested interest in obscuring knowledge of racism within a white supremacist society (Mills 2007, 2014) and thus a vested interest in promoting misuses of "racism," we can predict that enfranchised semantic drift will eventually corrupt the semantics of "racism" if left unchecked.

Enfranchised semantic drift presents an avenue for dominant groups to collectively practice what Pohlhaus (2012) calls willful hermeneutical ignorance. Willful hermeneutical ignorance occurs when dominant agents intentionally resist understanding or uptaking conceptual resources developed by marginalized groups for understanding oppression. When a whole community exercises its ability to control the semantics of its language in order to divest marginalized terminology such as "racist" of its emancipatory representational function, the community

effectively refuses to use or understand the word in a way that carries information about racism. This constitutes a kind of collective willful hermeneutical ignorance. It makes claims about real racism unintelligible and simultaneously verifies claims about "racism" made by the dominant group. The ability to interface only with some deference networks and fully reject others presents an opportunity to live in a kind of metasemantic echo chamber, in which the meanings of our words reflect the beliefs we share with those who are politically aligned with us.

A similar phenomenon also facilitates the creation of marginalized epistemic resources in the first place, the very resources dominant groups work to shut out. The practice of creating a mutually reinforcing deference network walled off from dominant linguistic practices is crucial for establishing and transmitting concepts and word meanings that function to illuminate and attack structures of oppression. Deferential networks of people of color and their allies are able to maintain a resistant use of the word "racist" by (a) refusing to let their usage patterns be swayed by or conform to the usages promoted by dominant linguistic networks and their deference magnets, and (b) allowing their own usage patterns to be shaped through deference to those who are seen to have the most epistemic authority in the domain under consideration.

To illustrate (b): members of the Black Lives Matter movement offer the refrain crafted by Representative Ayanna Pressley, "The people closest to the pain should be closest to the power." This refrain could also function as a guide to establishing deference magnets within resistant communities. Those who have experienced the most pain as a direct consequence of oppressive circumstances should be given higher degrees of deference concerning the correct usage of the terminology that functions to express or theorize about aspects of those experiences. Arguably, this deference practice is already enacted within networks of resistant populations such as the Black Lives Matter Global Network.

There can be subtle networks within resistant deferential networks. These subtle networks reflect differences of opinion about proper usages within marginalized communities. For example, there could be differences in opinion within BLM about the correct use of "racism." The semantic corruption model can capture these subtle patterns. It can identify subtle deference magnets within communities that are, broadly speaking, mutually deferential. For example, members of BLM will defer to one another much more strongly with regard to the correct usage of "racist" than they would defer to Donald Trump's uses of "racist," which are seen to be hijackings, yet within the BLM community there may be subtle differences in the strength of deference relations among community members that reflect the process of negotiating the meaning of the term within that community.

The existence of counterhegemonic subnetworks has been theorized in alternative ways by a number of scholars. According to Lugones (2006), for example,

we must see communication within "coalitions of the oppressed" as dependent on recognizing and navigating *unintelligibility*—the fact that shared interpretation of terms within resistant communities cannot be assumed. According to Lugones, the presupposition that communication within counterhegemonic subnetworks is "semiotically transparent"—that participants can easily understand one another's words—is a counterproductive fiction. Effective communication can only proceed through elaborate and complex measures, including storytelling and the patient sharing of memories to build connections and mutual comprehension; this is part of what Lugones calls "complex communication."

Lugones's perspective appears to conflict with the semantic corruption model on which members of linguistic subnetworks de facto use words with the same meanings. But the two approaches aren't really inconsistent, and seeing why helps us understand what the semantic corruption model entails. On the semantic corruption model, using a word with a shared meaning among members of a linguistic subcommunity does not entail that everyone in the subcommunity understands the word in the same way. It only means that members of the subcommunity are disposed to defer to one another in an attempt to correct and adjust their own understandings of the term. The process of actually adjusting conceptions might involve the kinds of patient and complex conversations Lugones recommends. On the semantic corruption model, the meaning of a word is an abstraction representing (roughly) what a community would eventually convince each of its members to accept. Complex communication might be part of the process whereby coalitions of the oppressed come to share an understanding of the meanings of crucial terms. Moreover, coalitions of the oppressed might not constitute the kinds of linguistic subnetworks that the semantic corruption model describes. Resistant coalitions are not founded on mutual linguistic deference but on shared interest in opposing oppression. So it could be that complex communication is part of transforming a political coalition into a resistant deferential network that uses crucial terms with shared meanings.

I have outlined a variety of ways in which linguistic hijacking can serve as a source of misinformation and active ignorance. I have also showed how the semantic corruption model can help us understand the metasemantic practices of resistant coalitions. In the next section, I raise an objection to the semantic corruption model. Then in section 4, I respond to this objection.

3. Semantic Data Objection

On the semantic corruption model, people who assert, "Affirmative action is a form of racism," do not refer to racism. Moreover, those who hear and accept this claim thereby defer to the community that uses "racism" to refer to mere discrimination and thereby consolidate their usage as referring to something other

than racism. These people do not form false beliefs about racism. They don't form beliefs about racism at all. They form true beliefs about the much less harmful—because nonsystematic—phenomenon of intentional race-based discrimination. Thus the semantic corruption model misses the fact that linguistic hijacking can involve *false* assertions that harmfully propagate *false* beliefs about racism.[16]

Recall the case of Richard Dawkins, who once tweeted that it was "racist" and "sexist" to call him a white man. It is quite natural to take him to have asserted that it was racist and sexist to call him a white man. Given background commitments from feminist and critical race theory, it is also quite natural to take what he asserted to be false and to reveal his false beliefs. Moreover, the false assertions are seen to spread false beliefs to others about the nature of racism and sexism. Insofar as these claims are accepted as semantic data, the semantic corruption model can be charged with getting the semantic data wrong.

Our first-order commitments about the truth or falsity of sentences invoking "racist" and "racism" are *data* that must be accounted for by an acceptable metasemantic theory. This data, from the perspective being illustrated here, includes the fact that "affirmative action is racist" is false, which is what our best theories of racism tell us. Our understanding of oppression properly informs our judgment about which sentences are true in the language we speak. These judgments in turn constrain which metasemantic theories we should find epistemically acceptable for that language. According to this objection, the semantic corruption model should be rejected because it fails to capture crucial semantic data. It incorrectly represents people who say, "Affirmative action is racist," as speaking truly.

This objection aligns with a venerable tradition in the philosophy of language. Kripke's (1972) famous argument against the description theory of reference for proper names is of the same form. Kripke argues that the description theory delivers the wrong semantics. It entails, for example, that a speaker who

---

[16] A related worry I have is the following. The semantic corruption model seems to degrade the value/harm of thinking and speaking truly/falsely in political discourse. For example, on the semantic corruption model, the harm of white people calling people of color "racist" has nothing to do with thinking or speaking falsely, since they speak truly (in their own white idiolect). A person can speak truly about "racism" even when they are ignorant about racism, so what is the value of speaking truly? But of course the existence of ambiguity does not mean that there is no value whatsoever in speaking truly. On the semantic corruption model, "racist" becomes ambiguous. It is still very valuable to speak truly when the word is used with its antiracist meaning. This value is part of what makes linguistic hijacking harmful. Thanks to an anonymous reviewer for suggesting this reply.

associates "Einstein" with the description "the inventor of the atomic bomb" would not refer to Einstein. The first-order semantic fact that "Einstein" refers to Einstein even when used by a confused speaker provides a decisive reason to reject the description theory. Likewise, when Iowa Representative Steve King claimed it would be "racist" to put Harriett Tubman on the twenty-dollar bill, what he said was false.[17] His use of "racist" denoted racism regardless of his confusion. This kind of first-order semantic fact provides epistemic force, analogous with Kripke's example, against the semantic corruption model.

4. Metalinguistic Negotiation and Semantic Corruption

I will now develop a response to this objection. The response is enabled by adopting a certain perspective on disagreement developed by Plunkett and Sundell (2013), according to which many conceptual disagreements are characterized by *metalinguistic negotiation*. A defender of the semantic corruption model can diffuse the objection raised in section 3 by understanding political disputes as a matter of metalinguistic negotiation. The semantic corruption model also has a natural affinity for this perspective, as we will see.

Metalinguistic negotiation occurs when parties to a disagreement use a shared term with different meanings. Each side then attempts to persuade the other to adopt their own preferred meaning. There is no disagreement over the truth of some disputed nonlinguistic proposition. The claims made by each party are true given the way each respectively uses relevant words.

Take one of Plunkett and Sundell's illustrations, a debate over whether a champion racehorse can be considered an "athlete." One party says the horse is an "athlete"; the other says it is "not an athlete." Plunkett and Sundell (2013, 16) claim there is "little reason to think that the dispute . . . concerns straightforward factual matters about the topic at hand." Both sides understand the attributes and accomplishments of the horse. Rather, the dispute must be about the proper way to apply the word "athlete," whether those attributes and accomplishments suffice for veridical application of the term. The speakers thus use the word differently. One speaker uses "athlete" with a meaning that includes racehorses, the other uses it with a different meaning that excludes them, and each speaks truly, even though they appear to contradict one another. Person A says, "The horse is an athlete," person B says, "The horse is not an athlete," and yet they both speak truly because "athlete" means different things in each sentence. Plunkett and Sundell note this explicitly: "each speaker literally expresses a true proposition given the concept they in fact express with their term" (17). To see this dispute as a metalinguistic

---

[17] See Nussbaum (2016).

negotiation is to see it as a scenario in which each side is attempting to convince the other to adopt their preferred meaning.

Now consider the case of "racist." We imagine one person, say Richard Dawkins, who insists it is "racist" to call him white. We find his interlocutor on Twitter insists it is not "racist" to call him white. Within the metalinguistic negotiation framework, both sides speak truly, expressing different concepts with the word "racist." Likewise when one person says, "Affirmative action is racist," and the other says, "Affirmative action is not racist," each says something true given what they mean by "racist." The debate is aimed at convincing the other to adopt one's own preferred meaning.

This is exactly what the semantic corruption model entails. Each side in the debate speaks truly, using the word "racist" to mean what their community uses "racist" to mean, and each side does its best to exert pressure on the other side to adopt its own preferred usage. Moreover, as I will develop further in section 5, the semantic corruption model provides a socially embedded conception of metalinguistic negotiation. It's not merely pressure exerted against one's conversational partner that matters, but rather pressure exerted against whole deferential networks.

According to Plunkett and Sundell, a metalinguistic disagreement need not be explicitly metalinguistic on its face. It can assume the surface form of first-order discourse. In other words, a dispute about the meaning of the word "racism" can sound like a first-order, metaphysical dispute about the nature of racism. The disputants may even take themselves to be having a first-order debate about racism. Yet they may in fact be speaking with different meanings of "racism" and thus, unbeknownst to them, they may be engaging in metalinguistic negotiation.

Plunkett and Sundell argue that because metalinguistic negotiation is possible and often tacit, we cannot infer that interlocutors engaged in an apparently first-order disagreement are using their words with a shared meaning. The fact that people seem to be arguing about racism (as opposed to arguing about the meaning of "racism") could be misleading. They could be engaging in a tacit metalinguistic negotiation.

Widespread tacit metalinguistic negotiation is further supported by a certain view of what it is to have a theoretical dispute. Plausibly, when two disputants disagree about the nature of racism, they are proposing different theories of racism. Then each is likely to use their term "racism" in a different way within their theory. Each distinctive theory of racism presents its own distinctive (but homophonic) theoretical term "racism." Different theories of racism entail differences in the theoretical role of the term "racism," and this difference in theoretical role arguably

constitutes a semantic difference between the homophonic theoretical terms.[18] A disagreement about racism is then necessarily at least partially metalinguistic in nature, because it involves proposing different theories of racism involving distinctive (but homophonic) theoretical terms.

The plausibility of widespread tacit metalinguistic negotiation undermines the force of the semantic data objection. If much of what appears to be first-order disagreement is in fact metalinguistic negotiation, then the semantic data are not what they seem to be. The semantic data objection relies on accepting the claim that, for example, Richard Dawkins spoke falsely when tweeting that it was "racist" to call him white, and treating this judgment as data for metasemantic theorizing. That this is a datum depends on the judgment that Dawkins and his detractors use the term "racist" with the same meaning. But if Plunkett and Sundell are correct, then the dispute between Dawkins and his detractors is plausibly metalinguistic in nature, even though it appears to be a first-order disagreement. If tacit metalinguistic negotiation is taking place, then the judgment that Dawkins spoke falsely is incorrect. Instead, he spoke truly in his own language in accord with the theory of racism he is trying to promote—just as the semantic corruption model predicts. The proponent of the semantic corruption model therefore has a way to diffuse the semantic data objection, namely by denying that the data are what they seem to be. The metalinguistic negotiation framework provides a plausible basis for doing just that.

Everyone (more or less) speaks truly in conceptual disputes whenever metalinguistic negotiations are happening. This is because, in a conceptual dispute characterized by metalinguistic negotiation, everyone's statements are true given the concept they are expressing and advocating for. This is equally the case on the semantic corruption model, at least insofar as the speaker's usage accords with the meaning supported by their deferential network.[19] We are still interested in arriving at a true theory of the phenomenon we are arguing about, but truth itself cannot be

---

[18] See Lewis (1970) for a defense of this view of theoretical terms.

[19] Note: this is not a consequence of Burge's semantic externalism as it is usually developed. Just because, e.g., two people disagree about what "arthritis" is doesn't entail they are expressing different concepts. The view that everyone speaks truly in their own language usually enters the dialectic as an entailment of internalist metasemantics. But the semantic corruption model does have this consequence, even though it is not an internalist metasemantics, under the circumstance where (i) each participant in the dispute defers to a distinctive deferential network, (ii) the two deferential networks use the word with different meanings, and (iii) the speaker is doing a good job of expressing the meaning of the word, given how it is used in their own community.

the guide insofar as both parties already speak truly. Rather, we are trying to settle what exactly we should be talking about. As a consequence of our conceptual disagreement, we do not agree exactly about what the target of the theory is.

While truth in a language can be manipulated by misuses of words, ontology cannot be manipulated in the same way. The existence and nature of racism is not altered by shifting deference patterns—these shifts only affect the semantic properties of words as used within the various linguistic communities. The ontology of racism is wholly independent of usage patterns. While truth can be stipulated, ontology cannot be. Ontology then carries the political significance that truth is often thought to have. Truth is also important, insofar as it represents a connection between representation and reality, but the relevance of having true beliefs must always be checked using one's ontology. Dawkins may express a true belief when he says it is "racist" to call him white, but the antiracist advocate can point out that his true belief is not relevantly connected with the phenomenon of racism, that which actually matters.

The ontology of a theory can feature in the criticism of an opposing theory. Those who believe, pace Dawkins, that racism necessarily involves structural oppression can criticize Dawkins's theory of racism for failing to capture this feature, even as we recognize that his claims are true as he uses the word "racism." That is, critical race theorists can use their own first-order theory of racism to criticize Dawkins's understanding, even though his sentences are judged to be true on the semantic corruption model. The ontological commitments of critical race theory's conception of racism—specifically, the essential character of racism as a form of systemic, oppressive power—plays a crucial role in making this argument against Dawkins.

On the semantic corruption model, knowledge of the real ontology of the universe is what matters. Truth is somewhat less important because it is partly conventional, relativized to linguistic subcommunities established by networks of deference relations. Ontology is not relativized to linguistic community, nor is it determined by convention. It is postulated to exist independently of linguistic conventions or social behavior, unlike truth and reference. Speaking and believing truly is rendered less important than knowing the correct ontology, since one can believe truly while avoiding having true beliefs that connect to the important aspects of the ontological reality. Yet truth is still important, in the sense that one needs a true theory that comports with the actual ontology of the universe in order to be politically on-track. A true theory remains necessary to political knowledge, even if its status as a sufficient condition has been dislodged.

5. Concluding Remarks: Metalinguistic Negotiation in a Social World

In their original paper, Plunkett and Sundell consider metalinguistic disputes as implicit arguments about word meanings carried on within a conversational context. The picture developed here suggests an expanded understanding of the phenomenon of metalinguistic negotiation, one that emphasizes dynamic change within a social fabric. On the semantic corruption model, metalinguistic negotiation involves attempting to alter patterns of deference. This expands the scope of metalinguistic negotiation. Arguments can be effective, in that they can have metasemantic effects, even when one's opponent doesn't budge in their own usage pattern.

Consider a public debate, perhaps on a news program, over what counts as "racism." The host uses "racism" to refer to mere race-based discrimination; his guest uses it to refer to racism. Even if the guest never concedes to the host's usage, the host can do work to reinforce oppressive deference networks among viewers, thus spreading his preferred semantics. Similar events can occur over new media as well as old. Indeed, such debates rage daily on Facebook pages across the Internet. These debates facilitate group metalinguistic negotiations (although in some contexts the term "negotiation" may be a misnomer). The semantic corruption model places metalinguistic negotiation into broader social contexts, representing the impacts of disagreements in terms of socially embedded actions and their effects on networks of deference.

Because political disputes are happening within a field of widely ranging social relations, metalinguistic negotiations carry a significance that goes beyond philosophical disputes understood within traditional philosophical dialectic. They have public import and proceed through public channels. They occur in the midst of power relations that affect how the meanings of words are spread or contained within those public channels. Their effects are not just a matter of refining theoretical understanding; rather, they are involved in political currents and social movements.

One crucial example of the way that metalinguistic negotiations are socially embedded is the phenomenon of enfranchised semantic drift discussed in section 2. The semantic corruption model allows us to see that the outcome of metalinguistic negotiations occurring on large scales will tend to favor those who carry more credibility within mainstream public spheres. This is due to a combination of factors, including the amplified effects of dominant deference magnets, who tend to enjoy greater and more widespread ascriptions of competence due to their social identities (as well as greater visibility), and the complimentary negative effects of conceptual competence injustice (Anderson 2017) that work against the strength of marginalized deference magnets using terminology in accord with marginalized perspectives. Without active intervention, we can predict that dominant semantic

deference networks will ultimately promote semantic properties that favor dominant groups and obscure the issues marginalized coalitions are trying to articulate.

The semantic corruption model thus provides a clear sense in which members of dominant groups have an ethical obligation[20] to 'speak the language of the oppressed' in order to actively push back against enfranchised semantic drift. The obligation arises from the fact that hijackings contribute to epistemic oppression (Dotson 2014). Persuasive but inaccurate uses of "racism" and "racist" erode the ability to use those terms as persuasively shared epistemic resources for producing and distributing knowledge of racism. This undermines the epistemic agency of people of color working to produce and distribute knowledge of racism in public, private, academic, legal, and political discourses. The obligation to speak the language of the oppressed includes the imperative that members of dominant groups (in this case, white people) must defer to members of marginalized groups in their uses of terms (in this case, "racist" and "racism") that describe, indicate, or define the nature of the oppression they face or the nature of their social identity.

The obligation also includes the imperative that members of dominant groups must actively resist linguistic hijacking of important terminology in both their own usage and in the usage of other dominantly situated agents. It is not enough for members of dominant groups to merely refrain from misusing politically significant terminology. A lack of action will tend toward semantic corruption. Active intervention on the part of dominant agents is required to resist this outcome. Therefore, dominant agents should be actively engaged in identifying and strategically intervening wherever they might be effective to forestall practices of linguistic hijacking in order to prevent the semantic corruption of marginalized epistemic resources.

Works Cited

Anderson, Derek E. 2017. "Conceptual Competence Injustice." *Social Epistemology* 31 (2): 210–223. https://doi.org/10.1080/02691728.2016.1241320.

Ball, Derek. 2020. "Metasemantic Ethics." *Ratio*. Published electronically February 4. https://doi.org/10.1111/rati.12256.

Bettcher, Talia Mae. 2009. "Trans Identities and First-Person Authority." In *You've Changed: Sex Reassignment and Personal Identity*, edited by Laurie J. Shrage, 89–120. Oxford: Oxford University Press.

---

[20] This obligation might fall under the field of study that Ball (2020) calls metasemantic ethics.

⸻. 2013. "Trans Women and the Meaning of 'Woman.'" In *Philosophy of Sex: Contemporary Readings*, 6th edition, edited by Nicholas Power, Raja Halwani, and Alan Soble, 233–250. Lanham, MD: Rowan & Littlefield.

Burge, Tyler. 1979. "Individualism and the Mental." *Midwest Studies in Philosophy* 4 (1): 73–122.

⸻. 1986. "Individualism and Psychology." *Philosophical Review* 95, no. 1 (January): 3–45. https://doi.org/10.2307/2185131.

Cho, Sumi, Kimberlé Williams Crenshaw, and Leslie McCall. 2013. "Toward a Field of Intersectionality Studies: Theory, Applications, and Praxis." *Signs: Journal of Women in Culture and Society* 38 (4): 785–810. https://doi.org/10.1086/669608.

Collins, Patricia Hill. 1986. "Learning from the Outsider Within: The Sociological Significance of Black Feminist Thought." *Social Problems* 33 (6): s14–s32. https://doi.org/10.2307/800672.

⸻. 1998. *Fighting Words*: *Black Women and the Search for Justice*. Minneapolis: University of Minnesota Press.

⸻. 2002. *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment*. 2nd edition. New York: Routledge.

⸻. 2015. "Sharpening Intersectionality's Critical Edge." Keynote address delivered at Social Theory Forum, University of Massachusetts, Boston, April.

Collins, Patricia Hill, and Sirma Bilge. 2016. *Intersectionality*. John Wiley & Sons.

Crenshaw, Kimberlé Williams. 1988. "Race, Reform, and Retrenchment: Transformation and Legitimation in Antidiscrimination Law." *Harvard Law Review* 101 (7): 1331–1387. https://doi.org/10.2307/1341398.

⸻. 1991. "Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color." *Stanford Law Review* 43 (6): 1241–1299. https://doi.org/10.2307/1229039.

⸻. 2018. "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory, and Antiracist Politics." In *Feminist Legal Theory*, edited by Katharine T. Bartlett and Rosanne Kennedy, 57–80. New York: Routledge.

Davis, Angela Y. 2011. *Women, Race & Class*. New York: Vintage.

Dotson, Kristie. 2012. "A Cautionary Tale: On Limiting Epistemic Oppression." *Frontiers: A Journal of Women Studies* 33 (1): 24–47. https://doi.org/10.5250/fronjwomestud.33.1.0024.

⸻. 2014. "Conceptualizing Epistemic Oppression." *Social Epistemology* 28 (2): 115–138. https://doi.org/10.1080/02691728.2013.782585.

Dretske, Fred I. 1991. *Explaining Behavior: Reasons in a World of Causes*. Cambridge, MA: MIT Press.

Fodor, Jerry A. 1990. *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.

Fricker, Miranda. 2007. *Epistemic injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Gotanda, Neil. 1991. "A Critique of 'Our Constitution is Color-Blind.'" *Stanford Law Review* 44 (1): 1–68. https://doi.org/10.2307/1228940.

Greenberg, Mark, and Gilbert Harman. 2008. "Conceptual Role Semantics." In *The Oxford Handbook of Philosophy of Language*, edited by Ernest Lepore and Barry C. Smith, 295–322. Oxford: Oxford University Press.

Hawthorne, John. 2007. "Craziness and Metasemantics." *Philosophical Review* 116 (3): 427–440. https://doi.org/10.1215/00318108-2007-004.

Henning, Tempest M. 2020. "Racial Methodological Microaggressions." In *Microaggressions and Philosophy*, edited by Lauren Freeman and Jeanine Weeks Schroer, 251–272. New York: Routledge.

hooks, bell. 1981. *Ain't I a Woman: Black Women and Feminism.* Boston, MA: South End Press.

Kennedy, Duncan. 1990. "A Cultural Pluralist Case for Affirmative Action in Legal Academia." *Duke Law Journal*, vol. 1990, no. 4 (Sept.): 705–757.

Kripke, Saul A. 1972. "Naming and Necessity." In *Semantics of Natural Language*, edited by Donald Davidson and Gilbert Harman, 253–355. Springer Netherlands.

Lawrence, Charles R., III. 1987. "The Id, the Ego, and Equal Protection: Reckoning with Unconscious Racism." *Stanford Law Review* 39 (2): 317–388. https://doi.org/10.2307/1228797.

Lewis, David 1969. *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

———. 1970. "How to Define Theoretical Terms." *Journal of Philosophy* 67 (13): 427–446.

Lorde, Audre. 2012. *Sister Outsider: Essays and Speeches*. Berkeley, CA: Crossing Press.

Lugones, Mariá. 2006. "On Complex Communication." *Hypatia* 21 (3): 75–85. https://doi.org/10.1111/j.1527-2001.2006.tb01114.x.

McKinnon, Rachel 2016. "Epistemic Injustice." *Philosophy Compass* 11 (8): 437–446. https://doi.org/10.1111/phc3.12336.

Millikan, Ruth Garrett. 1984. *Language, Thought, and Other Biological Categories: New Foundations for Realism*. Cambridge, MA: MIT Press.

Mills, Charles W. 2007. "White Ignorance." *Race and Epistemologies of Ignorance*, edited by Shannon Sullivan and Nancy Tuana. Albany: State University of New York Press.

———. 2014. *The Racial Contract*. Ithaca, NY: Cornell University Press.

Nussbaum, Matthew. 2016. "House GOP Dodges Vote to Block Harriet Tubman from $20 Bill." *Politico*, June 21. https://www.politico.com/story/2016/06/house-could-vote-to-block-harriet-tubman-from-20-bill-224637.

Peller, Gary. 1990. "Race Consciousness." *Duke Law Journal*, vol. 1990, no. 4 (Sept.): 758–847.

Plunkett, David, and Tim Sundell. 2013. "Disagreement and the Semantics of Normative and Evaluative Terms." *Philosophers' Imprint* 13 (3): 1–37. http://hdl.handle.net/2027/spo.3521354.0013.023.

Pohlhaus, Gaile, Jr. 2012. "Relational Knowing and Epistemic Injustice: Toward a Theory of Willful Hermeneutical Ignorance." *Hypatia* 27 (4): 715–735. https://doi.org/10.1111/j.1527-2001.2011.01222.x.

Rayo, Agustín. 2010. "A Metasemantic Account of Vagueness." *Cuts and Clouds: Vagueness, Its Nature, and Its Logic*, edited by Richard Dietz and Sebastiano Moruzzi, 23–45. Oxford: Oxford University Press.

Roberts, Dorothy E. 1991. "Punishing Drug Addicts Who Have Babies: Women of Color, Equality, and the Right of Privacy." *Harvard Law Review* 104 (7): 1419–1482.

Schiffer, Stephen R. 1987. *Remnants of Meaning*. Cambridge, MA: MIT Press.

Sider, Theodore. 2013. *Writing the Book of the World*. Oxford: Oxford University Press.

Simchen, Ori. 2017. *Semantics, Metasemantics, Aboutness*. Oxford: Oxford University Press.

Stalnaker, Robert C. 1997. "Reference and Necessity." In *A Companion to the Philosophy of Language*, edited by Bob Hale and Crispin Wright, 534–554. Malden, MA: Blackwell.

Sullivan, Shannon, and Nancy Tuana, eds. 2007. *Race and Epistemologies of Ignorance*. Albany: State University of New York Press.

Tomlinson, Barbara. 2013. "Colonizing Intersectionality: Replicating Racial Hierarchy in Feminist Academic Arguments." *Social Identities* 19 (2): 254–272. https://doi.org/10.1080/13504630.2013.789613

Williams, J. Robert. G. 2007. "Eligibility and Inscrutability." *Philosophical Review* 116 (3): 361–399. https://doi.org/10.1215/00318108-2007-002.

Williamson, Timothy. 2008. *The Philosophy of Philosophy*. Malden, MA: Wiley-Blackwell.

Yalcin, Seth. 2014. "Semantics and Metasemantics in the Context of Generative Grammar." In *Metasemantics: New Essays on the Foundations of Meaning*, edited by Alexis Burgess and Brett Sherman, 17–54. Oxford: Oxford University Press.

Dr. DEREK ANDERSON is a lecturer at Boston University. He works in a range of areas including feminist epistemology, philosophical logic, metasemantics, intersectionality studies, and the history of analytic philosophy. He's currently working on a book entitled *Metasemantics and Intersectionality in the Misinformation Age*.