

Racialized Discourse at the Intersection of Meaning, Mind, and Metaphysics

by

Patrick O'Donnell

A dissertation submitted to Johns Hopkins University in conformity with the requirements

for the degree of Doctor of Philosophy

Baltimore, Maryland

February 2019

Abstract

Racialized discourse is language that transmits potentially harmful representations of racial groups. It is also a tool for maintaining status quo racial hierarchies. A theory of racialized discourse should describe the form and content of racialized representations, explain how they are transmitted in communication, and explain how their distribution plays a role in sustaining racial hierarchies. I meet these desiderata via an original account of the semantics of racially stereotypical generics (e.g. “Blacks are criminal,” “Muslims are terrorists,” “Immigrants are violent”) and racialized terms deployed in the context of political discourse (e.g. “thug,” “terrorist,” “immigrant,” “criminal,” “welfare”). The core semantic hypothesis is that the standards for the *use* and *meaning* of racialized vocabulary shift depending on the racial presentation of the individuals and groups described by that vocabulary. This shows that racial discrimination sometimes has a linguistic basis. Next, drawing on an interdisciplinary set of tools offered by philosophy of language, linguistics, developmental and social psychology, political science, and social ontology, I show that these types of language i) essentialize racial groups, ii) indirectly increase tolerance for social hierarchy, and iii) play a role in maintaining racial stratification.

Committee:

Steven Gross (advisor)

Christopher Lebron (second reader)

Justin Bledin (third reader)

Lester Spence (external reader, political science, Johns Hopkins University)

Bernhard Nickel (external reader, philosophy, Harvard University)

For my parents, Angela Alaimo O'Donnell and Brennan Patrick O'Donnell, first and best teachers, and learned doctors in their own right.

Preface and Acknowledgements

Saying racist things gets people to believe racist things. When people believe racist things, they act in racist ways. When people act in racist ways, they sustain racial hierarchies. Such, in any case, is the intuitive story of the relationship of language to racial domination. Thinkers primarily interested in racial domination more or less assume that something like the intuitive picture is right, but they don't usually concern themselves with piecing the different components of the story together. Thinkers primarily interested in language and communication think that something like the intuitive story *may* be right, but that's only if they think about the intuitive story at all. In this dissertation, I strive to have something to say to both of these camps. In short: the intuitive story is right, and in explaining *why* it's right, we'll learn something interesting and novel about the relationship between racism, language, and social practices. Hopefully this is a benefit to anyone interested in race, racism, and language.

My interest in the intersection of race and language has a long history. I grew up in Baltimore city in the 1990's. On a local level, White flight, the crack epidemic, an unprecedented spike in violent crime, increasing de-industrialization, and the rise of an increasingly punitive carceral apparatus were the most visible symptoms of and responses to deeply entrenched racial and class-based inequities and tensions. On a national level, young Black males, single Black mothers, and Muslims were fast becoming the avatars of a nation's racial, economic, and religious anxieties. Twice in this decade, Bill Clinton coasted to the Oval Office on the wheels of a draconian crime bill, a campaign to "end welfare as we know it," and a selectively interventionist foreign policy.

As the privileged (White) son of middle-class academics, I was by no means directly victimized by this social, racial, and economic turmoil. Yet geographical, social, and political stratification based on race has a tendency to make racial difference the elephant in the room. Early

on I learned that when Baltimoreans--White or Black-- were opining about crime rates, or sexuality, or after-school programs, or complaining about taxes, or bemoaning the state of public education, or criticizing religious institutions, or bashing Democrats and Republicans, or advocating for more city housing and public transportation, or talking about practically anything, they were in some sense talking about race, whether they knew it or not. There was even a vocabulary for it, a subtle repurposing of familiar words floating just beneath the surface of everyday idle discourse, in which people traded ideas about racial differences. When I got older, I found out that when Baltimoreans talked about race in this way, they were engaging in some small-scale version of politics. And when I got older than that, I figured that if this small-scale politics had anything to do with larger-scale forms of racial domination, I wanted to know how they were related.

This dissertation represents one way of working out this long-standing interest. It is a progress report, representing my current stated view of at least one possible relationship between racialized discourse and racial domination. In many respects, the view is incomplete. It is undoubtedly wrong, at least in some of its details. Nevertheless, it is my hope that its incompleteness invites collaboration and continuation, and that where it is wrong, it is wrong in a way that is interesting and useful.

When I started my doctoral studies at Hopkins in 2011, there was no one on the permanent faculty who worked on race from a philosophical angle. (This was before the 2017 arrival of Chris Lebron, whose service on my committee I gratefully acknowledge.) During this time I was largely an autodidact, reading everything I could on race and racialization in the American context with only sporadic direction. Two people were instrumental in lending some much-needed guidance. First, Laura Papish, then a visiting assistant professor of philosophy at JHU, first pointed me toward

important resources in the philosophy of race, and it was largely through her intervention that I even became aware of there being any such sub-discipline in philosophy. Second, Lester Spence's 2013 course "Race and the Neoliberal Turn" was a formative influence on my thinking about the relationship between "micro" and "macro" manifestations of racism, and his course pointed me to invaluable theoretical and empirical resources that broadened my view of the sorts of interesting philosophical questions one could ask about race and racism. I am also grateful to Lester for serving on my committee as an external reader.

Steven Gross has been an ideal advisor throughout the course of this project. Even more impressive than the sheer breadth and depth of his knowledge of philosophy and its relationship to the empirical sciences is his meta-philosophical orientation, which seems to find something interesting, worth thinking about, and worth questioning in practically everything. Steven's willingness and ability to think through thorny issues in ways that are rigorous, creative, and sensitive has not only deeply shaped this project, but has also served as a regulative ideal for me. In 2012, Steven led a seminar on contextualism in the philosophy of language, which provided me with much of the background I needed to start thinking about the problem of racialized discourse in earnest. He also provided feedback (in a collaborative, collegial spirit) on earlier versions of most chapters in written form, as well as in bi-weekly "lab meetings" during which many of the basic ideas of this dissertation were presented.

While we approach the topic from different angles, Bernhard Nickel's work is largely responsible for my interest in genericity and the logic of conversation. At different points, Bernhard supplied incisive feedback on the content of Chapters 2 and 4, which greatly improved the structure of the argument. I also thank Bernhard for graciously agreeing to serve as an external reader on

relatively short notice. Thanks also to Justin Bledin and Dean Moyar, who each provided helpful feedback when this project was in its earliest stages, and raised formidable problems that challenged me to revise and reconsider my approach to these issues. Thanks especially to Justin for serving as a reader for this dissertation.

Thanks also to friends and colleagues at Hopkins over the years, who provided stimulating conversation relevant to this work and/or lent a blessed reprieve from thinking about it: Jonathon Hricko, John Waterman, Sandy Koullas, David Jacobs, Danny Schwartz, Josh McBee, Karen Yan, Tom Wilk, John Brandau, and David Lindeman. Thanks are also due to Alicia Myers and Veronica Feldkircher-Reed, who were invaluable throughout my time in the philosophy department. Thanks to the Johns Hopkins Counseling Center for clinical support over the years when it became difficult to reconcile my station and its duties with the project of maintaining a *mens sana*.

Thanks to Daniel Estrada, modern-day Zarathustra, in all likelihood the most important true philosopher of his generation, who has had a massive influence on the way I think about philosophy, methodology, pedagogy, and the universe in general. For better *and* for worse, Dan's fingerprints are all over this and anything else I'm likely to do in philosophy.

Thanks to my many friends and acquaintances in scenes and places far from the ivory tower. You have kept me grounded and cultivated in me a healthy ambivalence and judicious skepticism about at least some of the things that the academy seems to value. Perhaps the best way I can thank you all is by not dragging you each by name through the acknowledgments section of an academic dissertation. Thanks to horror for perspectives cosmic and comic. Thanks to heavy metal for disharmonic convergence and scorched-earth frisson.

Thanks to my parents, Angela and Brennan O'Donnell, for giving my life direction, for encouraging me to cultivate my talents, for instilling in me worthwhile values, lending emotional and often financial support, and generally being members of that relatively small class of people who are actually cut out for this parenting thing. None of us ask to be born, and some of us might even be worse off for it. But if I am numerically identical to some individual that had to be born anyway, I am unbelievably lucky and privileged to be one of the individuals born to this particular pair. I would not have been able to finish this program (or likely even start it) without their support and love, and so this work is dedicated to them. Thanks are also due to the other two individuals born to this pair, my brothers Charlie and Will, who have supported me through a number of difficult stretches during my doctoral career, and whose own (much harder) work in education lends me both inspiration and perspective.

Finally, thank you to Theresa, rose of my heart, supportive partner, scintillating educator, and inspiring interlocutor, who shows me anew each day that when you really get down to it, love is the only thing worth living and dying for in this life, and everything else can get bent.

Table of Contents

List of Figures.....	xiii
-----------------------------	-------------

Chapter 1. Toward a Theory of Racialized Discourse.....1

1. Racialized discourse: The basic idea.....	4
1.1 Racialization.....	4
1.2 Discourse.....	6
1.3 The contribution.....	8
2. Racialized discourse and standard-shifting.....	8
2.1 Use-standards.....	9
2.2 Meaning-standards.....	12
2.3 Conduct-standards.....	16
3. Racialized discourse and social perspectives.....	20
4. Structure of the dissertation.....	24

Chapter 2. Generics, Race, and Social Perspectives.....35

1. What are generics?.....	36
2. Leslie's weak semantics: Type A and Type B generics.....	39
2.1 Genericity and primitive generalizations.....	39
2.2 A weak semantics for generics.....	41
3. Introducing Type C generics.....	47
3.1 Socially perspectival predicates.....	47
4. Leslie reconsidered.....	53
4.1 Contra Leslie 1: Perspectives and semantics.....	54
4.2 Contra Leslie 2: Racialized interpretations and weak semantics.....	57
4.3 Backing things up.....	61
5. Socially perspectival semantics.....	61
5.1 Proof of concept: Baseball.....	62
5.2 Racial perspectives.....	68
5.3 A characterization of perspectives.....	72
5.4 Perspectives and linguistic meaning.....	73
6. The semantics of (some) Type C generics.....	76
7. Conclusion.....	83

Chapter 3. When Code Words Aren't Coded.....	86
1. The standard framing.....	90
1.1 The communication theory.....	90
1.2 The semantic theory.....	93
2. Limitations of the standard framing.....	95
2.1 Examining Mendelberg's assumptions.....	97
3. Candid racial communication.....	106
3.1 Two examples of candid racial communication.....	107
3.2 The normative structure of candid racial communication.....	108
4. Racialized terms and referential modification.....	112
4.1 Referential restriction.....	115
4.2 Referential expansion.....	118
5. Community-specific speech and referential content.....	119
6. Conclusion.....	127
Chapter 4. The (Meta)Semantics of Racialized Terms.....	128
1. What are racialized terms (again)?.....	131
1.1 A negative definition of racialized terms.....	132
2. What should a semantics for racialized terms explain?	135
2.1 Threshold problem.....	135
2.2 Defeasibility problem.....	136
3. Semantics and pragmatics.....	137
4. Metasemantics: context-sensitivity, meaning underdetermination, social perspectives...142	
4.1 Context-sensitivity.....	143
4.2 Meaning underdetermination.....	148
4.3 Social perspectives revisited.....	156
4.4 Combining the metasemantic hypotheses.....	162
5. A first-order semantics for some racialized terms.....	163
5.1 Genericity again.....	164
6. Racialized terms as inherent generics.....	174
7. Conclusion: Thresholds and defeasibility.....	179
7.1 The threshold problem.....	180
7.2 The defeasibility problem.....	181
7.2.1 Defeasibility due to shifts in generality.....	181
7.2.2 Defeasibility due to meaning-underdetermination.....	182

Chapter 5. Essentializing and Moralizing Racialized Conduct.....184

1. Essentialism and racialized categories.....	188
1.1 Essentializing racialized categories.....	191
2. Immigrants, terrorists, and welfare recipients.....	193
2.1 Immigrants.....	194
2.2 Terrorists.....	196
2.3 Welfare recipients.....	199
2.4 Racial essences.....	202
3. Essence, explanation, and intergroup attribution.....	202
3.1 A brief history of attribution errors.....	204
3.2 Essentialism and attribution.....	208
4. Is essentialist thinking compatible with moralistic thinking?.....	214
4.1 Shifty essences and essentialist explanation.....	215
5. Moralistic attribution.....	221
5.1 Muslim and non-Muslim “terrorism”.....	223
5.2 White and Black welfare recipients.....	226
5.3 White and Black athleticism.....	227
5.4 MAE as a working hypothesis.....	230
6. Conclusion.....	231

Chapter 6. Racialized Discourse and the Flow of Information.....233

1. Generics and essentialism.....	236
1.1 Generics and the social transmission of essentialist belief.....	237
2. Generics and the common ground.....	243
3. Haslanger on generics and pragmatics.....	248
4. Racialized discourse as both essentializing and legitimizing.....	255
4.1 Status quo bias and system justification.....	257
4.2 The psycho-pragmatic function of generic racialized discourse.....	261
5. Conclusion: Where we’ve been, and what’s next.....	264
5.1 The argument so far.....	265

Chapter 7. Constructing Racialized Categories.....268

1. Racializing criminality: a historical example.....	271
2. Racialization, category membership, and tracking.....	276
2.1 The tracking relation.....	279
2.2 Multiple realizability.....	284
2.3 Multiple realizability and social construction.....	287
2.4 What SRCs are.....	290

3.	Conferralism and racialization.....	292
3.1	Conferralism and social categories.....	293
3.2	Conferralism and SRCs.....	297
4.	Conclusion: Racialization as a metaphysical and linguistic phenomenon.....	300
Appendix. Against Semantic-Level Perspectivalism.....		304
1.	Semantic-level perspectivalism.....	304
2.	Different types of perspectives.....	306
3.	Three problems for semantic-level perspectivalism.....	311
3.1	Is an RP parameter necessary?.....	311
3.2	Lack of evidence from syntax.....	311
3.3	The Binding Criterion.....	313
Bibliography.....		317
Curriculum Vitae.....		337

List of Figures

Figure 1: SPPs and Racialized Terms in the Space of Linguistic Meaning.....	48
Figure 2: The Strike Zone.....	64
Figure 3: Similarities and Differences Among Types of Racialized Discourse.....	134
Figure 4: I-level and S-level Predicates.....	166
Figure 5: Kratzer’s Semantics for S-level and I-level Predicates.....	168
Figure 6: Chierchia’s Semantics for S-level and I-level Predicates.....	170

Chapter 1. Toward a Theory of Racialized Discourse

To speak means being able to use a certain syntax and possessing the morphology of such and such a language, but it means above all assuming a culture and bearing the weight of a civilization.

--Frantz Fanon, *Black Skin, White Masks*, 1958

On January 19, 2014, during the National Football League's playoff season, the Seattle Seahawks defeated the San Francisco 49ers in a close game. Richard Sherman, a cornerback for the Seahawks, played a decisive role in his team's success, and was the subject of a now-famous postgame live television interview. Sherman, who is Black with long dreadlocks, was visibly animated, energized by his team's important victory. He yelled hoarsely into the camera. He ostentatiously flexed his muscles. He disrespected opposing San Francisco receiver Michael Crabtree. In short, he was boisterous, loud, and unsportsmanlike. However, many members of the mainstream sports media and millions of viewers had a far stronger reaction: in addition to being boisterous and unsportsmanlike, Sherman was a "thug." According to one media analysis, Sherman was labelled a "thug" over 600 times on broadcast television alone within a single 24-hour news cycle (Wagner 2014). Additionally, the "thug" label was used pejoratively to describe Sherman in countless other instances scattered across social media platforms such as Facebook and Twitter.¹

Why did so many viewers and commentators converge on this particular term to describe Sherman and his conduct? Even if Sherman was indeed acting in an impulsive and unsportsmanlike manner, there are a variety of epithets to hurl at him: perhaps the high-brow among us would call him a "braggart," while the rest of us would rest content with calling him an "asshole." So why "thug"? If the public discourse surrounding the interview is any indication, Sherman's detractors

¹ Of course, more explicitly racist invective involving slurs and racist imagery was also heaped upon Sherman across social media networks. See Howard (2014) for more background on Sherman and specific information on these more explicit sorts of reaction.

meant to ascribe traits such as aggressiveness, poor upbringing, ignorance, low intelligence, arrogance, and other negative psychological and moral traits. Perhaps “braggart” just wouldn’t convey all that. Yet many also hypothesized that it was Sherman’s *race* that explained why “thug” jumped so readily to brains, tongues, and fingertips. According to this theory, Sherman wasn’t *just* arrogant and loud, he was also arrogant, loud, and *Black*. Calling him a “thug” was a judgment about his conduct, to be sure, but that judgment was also intimately bound up with race-based evaluations. In fact, the author of the aforementioned media analysis opined that

The word ‘thug’ has been used so many times by the same sort of people about the same sort of thing that it’s no longer even accurate to call it code—it’s really more of a shorthand. It means a Black guy who makes white folks a little more uncomfortable than they prefer. (Wagner, 1/21/14)

Many pointed out that White athletes who act in boisterous and unsportsmanlike ways, while frequently sanctioned by public opinion, are not generally called “thugs.” Sherman himself drew attention to this double standard in a starker way when he went so far as to suggest that “thug” “seems like an accepted way of calling somebody the N-word now” (Wilson 2014). On the other hand, many stood by their description of Sherman as a “thug,” and vociferously claimed that their use was neither racist nor affected by any Sherman’s racial presentation. Others bemoaned a seeming double standard which permitted Blacks to refer to themselves as thugs (especially in the context of hip-hop) but did not permit Whites to refer to Blacks in this manner.

Over a year later, another high-profile negotiation over the uses and abuses of “thug” surfaced. In the aftermath of widespread protests against police brutality which became violent on April 27, 2015 in Baltimore, MD, mayor Stephanie Rawlings-Blake (a Black woman) continually used the word “thug” to condemn the behavior of the rioters, a group which consisted almost entirely of

young Black males. Her rhetoric was widely echoed by media coverage of the protests and by then-President Barack Obama as well. In an interview on CNN the next day, (Black) Baltimore city councilman Carl Stokes objected to (White) reporter Erin Burnett's suggestion that "thug" was an appropriate term to use in characterizing the behavior of the rioters. After Stokes objected that it was misleading to describe angry Black teens as "thugs," Burnett followed up:

BURNETT: But how does that justify what they did? I mean that's a sense of right and wrong. They know it's wrong to steal and burn down a CVS and an old person's home, I mean, come on.

STOKES: Come on? So calling them thugs? Just call them [slur for Blacks.] Just call them [slur for Blacks]. No. We don't have to call them by names such as that. We don't have to do that. (Levs 2015)

These anecdotes suggest that the project of determining just who is and who is not a "thug" has high stakes. For instance, perhaps both Sherman and Stokes are right to worry that the *use* of the term both presupposes and is a causal effect of a long, unjust history in which nearly every aspect of the conduct of Black males has been more heavily scrutinized, policed, and sanctioned than that of the conduct of Whites. There is the possibility that the term has a racist *meaning*: both Sherman and Stokes are worried that "thug" is a vehicle for transmitting racist and dehumanizing ideas under the facade of social acceptability. Finally, there is the worry that the term has unjust social *effects*: it singles out certain agents as particularly worthy of repressive treatment: perhaps Stokes is justified in worrying that the term further marginalizes young Black males by portraying them as irredeemable predators rather than misguided youths.

If any of these worries are well-founded, it seems highly consequential for us, whether we are slugging it out in the linguistic trenches or surveying the terrain from the theorist's perch, to understand what work terms like “thug” are doing in the context of linguistic, social, and racial practices more broadly. What do terms like “thug” mean? What is their role in linguistic practice and communication? And how do they organize our racial reality?

1. Racialized Discourse: The Basic Idea

“Thug” is an example of what I will call *racialized discourse*. The goal of a theory of racialized discourse is twofold. First, it must explain how language manages to transmit racialized representations in everyday communication. Second, it must explain how those processes of representation-distribution concretely affect the socio-racial environment. In meeting the first goal, I argue that racialized representations in both thought and talk have *race-sensitive* meaning properties. That is, certain core semantic and pragmatic properties of these representations “shift” according to the racial presentation of the individuals, groups, institutions, or practices that they purport to represent. A major goal of the dissertation is thus to offer an account of what these “shifts” are, and when, how, and why they occur. In meeting the second goal, I argue that racialized discourse has the potential to shape *dominant understandings* of how the social-racial world is constituted in ways that are both politically and epistemically disabling. As such, I present a theory of what racialized discourse communicates among interlocutors, and how the content of what is communicated shapes the social world.

1.1 Racialization

The term “racialized discourse” is intended to capture a type of vocabulary or region of linguistic practice. I take racialized discourse to be analogous to more traditional philosophical discourses like “epistemic discourse,” “logical discourse,” “moral discourse,” and the like. While each of these discourses have concepts which are germane to them (such as “entailment” in the case of logic or “right action” in the case of moral discourse), these concepts do not “belong” to these discourses by virtue of sharing some deep semantic or logical feature. Rather, they belong to these discourses because they are useful for doing whatever the discourse’s practitioners think that they are up to in using, say, moral vocabulary, and in general they don’t concern themselves with whether there is some non-obvious *mark* which determines once and for all *why* all the different moral concepts belong under the umbrella of moral discourse.² I assume the same goes for racialized discourse in what follows-- we should not expect all the concepts which can plausibly be considered “racialized” to have much in common other than what they *do*. If there is anything that all the bits of racialized discourse have in common, it is how they behave in the context of linguistic and social practices.

So what does racialized discourse do? First and foremost, it *racializes*-- it assigns racial significances to bodies, practices, conduct, and institutions. What it means to “assign racial significances” is a vexed question. Yet we can make do with some examples. For instance, terms such as “White,” “Black,” “Asian,” “Native American,” and “Latino,” in the contemporary United States, reflect a widely held belief that human bodies have both obvious and non-obvious

² This goes for naive practitioners as well as sophisticates. While thinkers in the abstruse reaches of iterated branches of inquiry such as (meta)metasemantics or (meta)metaethics might devote a good deal of time trying to determine the mark which individuates (meta)semantic or (meta)ethical vocabulary, most of us are happy to take for granted that some concepts are more central to the disciplinary practice, and some are more peripheral.

characteristics that make them appropriately sortable into a variety of racial types.³ Cultural practices which are closely associated with specific sorts of racialized bodies might also be racialized themselves-- for instance, hip-hop might be racialized as Black, and country clubs might be racialized as White.

Dominant cultural understandings of social practices and categories can also reflect the effects of racialization. For instance, crime is racialized to the extent that Blacks and other non-Whites have long been perceived to be especially disposed towards criminality (Muhammad 2012). Similarly, punishment and incarceration are racialized institutions, insofar as they are disproportionately applied to Blacks and other non-Whites, and insofar as these institutions justify their operation partially by appeal to fears and anxieties about the supposedly ineradicable non-White tendency to violence and predation (Wacquant 2000, Alexander 2012). Immigration is racialized to the extent that the “immigrants” in question are usually perceived to be various shades of non-White, and terrorism is racialized (or at least indexed toward a specific religion) to the extent that Muslims (and especially Arab and African Muslims) are perceived to be paradigmatic examples of “terrorists.” Racialized discourse is a central mechanism of racialization, but by no means the only one.

1.2 *Discourse*

³ Americans take these races to be defined by some admixture of both biological and cultural features, and the data on what races *are* in common sense conception points in a number of different directions (Condit et al 2004, Jayaratne et al 2009, Glasgow 2009.) Luckily we need not answer the question of what races are in order to understand how racialization occurs. We return to the question of how racialized social categories are represented in common sense conception in Chapter 5.

There are two ways to think about the “discourse” in racialized discourse. On the one hand, discourse is encoded in the familiar vehicles of linguistic communication, talk and text. Call this *semantic* sense of racialized discourse. Understanding discourse in this sense requires us to explore how speakers and hearers transmit racialized representation via communication. As such, it requires an investigation into the *semantics* and *pragmatics* of racialized discourse (“what it means”) as well as an investigation into what sorts of *mental representations* of racialized categories are implicated in the interpretation and production of such discourse.

The second way of thinking about discourse is as a tool for shaping the social world in particular ways. This requires thinking of discourse as having the potential to *construct* the world as well as describe it. Call this the *social-metaphysical* sense of racialized discourse. Understanding discourse in this sense will require us to explore what sorts of effects the distribution of racialized representations have on the social world. It also requires us to provide some account of how racialized discourse can be cultivated in the pursuit of broader social and political projects.

In this dissertation, I will be concerned most centrally with racialized discourse which is *negatively valenced* on the semantic level, and which can reasonably be expected to shore up forms of *racial domination* on the social-metaphysical level. This is a theoretical choice of mine, since my theory leaves open the possibility that racialized discourse could be positively valenced in ways that can be put to liberatory ends. In this sense my understanding of racialized discourse is akin to Toni Morrison’s conception of “race talk,” which she defines as “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.” (Morrison 1993) The goal of the dissertation is to make explicit

the properties of the “signs and symbols” within the ambit of racialized discourse and to explain how they can be the sort of things that racially subordinate groups.

1.3 The Contribution

The dissertation provides an original theory of the meaning, use, and function of racialized discourse which explains the relevance of its properly *semantic* domain to its properly *social-metaphysical* domain. “Semantic” racialized discourse is a medium for the social transmission of racialized representations. “Social-Metaphysical” discourse is a tool for constructing racial realities. Each domain needs the other in order to properly function. The transmission of racialized representations through language creates racial realities. Racial reality, carved as it is in the image of racialized discourse, then repays language in kind, providing the conditions under which racialized discourse can seem truthful, reasonable, and normal.

We need two core concepts in order to understand how racialized representations are distributed and how those representations shape the social world. The first is the notion of a *standard-shift*. The second is that of a *social perspective*.

2. Racialized Discourse and Standard-Shifting

My first core claim is that *racialized discourse fixes and exploits different standards, norms, and expectations for agents, behaviors, and practices in a way that is systematically linked to how those agents, behaviors, and practices are racially interpreted*. The specific sense in which racialized discourse “fixes and exploits” such standards differs across the semantic and social-metaphysical domains. Within the semantic domain, the standards which determine the *use* and the standards which determine the *meaning* of

terms in a language differ according to what racialized agents and institutions are being described by those terms. Call these *use-standards* and *meaning-standards*. Within the social-metaphysical domain, the normative standards which determine the relative acceptability of a given “move” in the social domain shift according to which racialized individuals are performing those moves. Call these *conduct-standards*. The goal of a theory of racialized discourse is to provide a theory of how all three of these standards shift “in the face of race.”⁴

2.1 Use-standards

Use-standards determine the conditions under which the use of a term or the staking of a claim is *licensed*. We can understand these licenses in various ways. Perhaps one is licensed to use term T to describe x just in case T is *true of x*, or if T is *assertable*, or if T is otherwise *permissible* at a context. The point about use-standards in the case of racialized discourse is simply that just what sorts of terms it seems appropriate to use shift with the perceived racial presentation of the individuals under discussion. Thus the notion of a shift in use-standards is ultimately a modal notion. The basic idea is that, for a given description, what counts as an acceptable description of racial group R would *not* count as an acceptable description of racial group R’.

Race-related shifts in use-standards: A description D reflects a race-related shift in use-standards when D is generally deemed acceptable when applied to members or subgroups within racial group R and unacceptable (or less acceptable) when applied to members or subgroups within a salient racial group R’.

⁴ This phrase, which I use throughout the dissertation, is due to Lebron (2013), who provides a theory of how observed standards governing democratic deliberation shift or are abandoned when race (particularly Blackness) becomes salient in these deliberations.

Obviously this isn't an exhaustive characterization, since there may be many senses in which a description can be "acceptable" (e.g. D is "acceptable" iff D is a *true* description, or an *assertible* description, or a *fair* description, etc.) The general idea is that given a background conception of what "acceptability" amounts to, race-related shifts in use-standards are in effect when those standards seem to recommend that D is *especially* appropriate for a given racial group R, as opposed to any number of salient groups R'.

A few examples might clarify what I have in mind. In a study conducted in the wake of Hurricane Katrina, which decimated New Orleans and disproportionately affected poor Black neighborhoods, Sommers et al (2005) found that Blacks were much more likely to be described in media reports as "refugees" who were engaging in "looting" in the wake of Hurricane Katrina, as opposed to their White counterparts, who were more likely to be described as "evacuees" who were "finding food." In an analysis of over 1500 articles of the aftermath, Nunberg (2005) found that while "evacuee" was a far more common term than "refugee" in describing the citizens of New Orleans in general, the term "refugee" was used approximately twice as often compared to the word "evacuee" when either term occurred within 10 words of the terms "poor" or "Black." While we need not suppose that this was an intentional choice, this is strong evidence that a race-related shift in use-standards is in effect: "refugee" is simply less acceptable as a description of Whites than it is of Blacks.

These findings are consistent with van Dijk's (2000) observation that racial stereotypes, motivated reasoning, and biased perceptions can affect the "lexical choices" made in describing events and social categories.

The new racism... avoids explicitly racist labels, and uses negative words to describe the

properties or actions of immigrants and minorities (for instance, “illegal”.) Special code-words (such as “welfare mothers”) may be used, and the readers are able to interpret these words in terms of minorities and the problems attributed to them. And it needs no further argument that attitudes about groups and opinions about specific events may influence the “lexical choice” of such words as “riot” on the one hand, or “urban unrest,” “disturbance,” or “uprising” on the other hand, as is the case for the classic example of “terrorist” vs. “rebel” vs. “freedom fighter.” (van Dijk 2000: 39)

As a good deal of survey data makes clear, contemporary social and political pressures often play a role in determining what sorts of race-related shifts in use-standards are operative at a given context. For instance, Americans’ go-to descriptions of other nations often shift with military and economic policy. In many of these cases, descriptions which are taken to be especially applicable to certain ethno-national characters change dramatically:

Every so often [Gallup polls] ask their respondents to select from a list of adjectives the ten which best describe members of other nations...Back in 1942, Germany and Japan were [America’s] bitter enemies, and Russia was our ally; and in 1942, among the first five adjectives chosen to characterize both the Germans and the Japanese were: “warlike,” “treacherous,” and “cruel.” None of these appeared in the list for the Russians at the time. In 1966, when Gallup surveyed responses to mainland China, the Chinese were seen as “warlike,” “treacherous,” and, being Orientals, “sly.” After President Nixon’s visit to China, however, these adjectives disappeared about the Chinese, and they [were] characterized as “hard-working,” “intelligent,” “artistic,” “progressive,” and “practical.” (Frank 1982, quoted in Smith 2011: 49)

Finally, race-related shifts in use-standards have also been discovered in computer science, via projects to debias machine classificatory algorithms. Rudiger et al (2017) find that a major human-generated corpus used for training up artificial natural language processing mechanisms contained a significant number of stereotypical racial semantic associations. For instance, when crowdworkers were asked to contribute sentences that were contradictory to, entailed by, or compatible with respect to a given target sentence, stereotypical racial and gender associations were likely to emerge in the user-generated sentences when the target sentences invoked racial or

gendered categories (e.g. “Muslim”/”terrorist,” “woman/hairdresser/receptionist,” “woman”/”slutty,” “Black man”/”jail,” etc.) Assuming that such stereotypes form the basis of further generalizations and are replicated in the normal functioning of the algorithm (Crawford 2016, Angwin et al 2016), this suggests that use-shifts could be characteristic of “artificial” as well as “human” language use.

In this project, I largely take for granted that shifts in use-standards are characteristic of racialized discourse. Since much of the evidence for race-specific shifts in use-standards is corpus-based, and has to do with the statistical frequency of when certain “lexical choices” rather than others are made in the face of race, there is little reason to doubt that racial presentation is a significant determinant of what lexical choices will be made. A more radical claim to make about racialized discourse in the semantic domain would be that the *meaning-standards* associated with a term shift systematically depending on which racialized individuals are being described by those terms.

2.2. *Meaning-Standards*

To say that meaning-standards shift in the face of race is to claim that it is possible for a description D to have different *contents* depending on the racial presentation of the individual or group D is describing at a context. Like the claim about use-standards, the notion of a shift in meaning-standards is a modal notion. The basic idea is that, for a given description D, D has meaning M when it describes racial group R at a context, but *would* have some other meaning M' if it were describing racial group R' at a context.

Race-related shifts in meaning-standards: A description D reflects a race-related shift in meaning-standards if D has meaning M when applied to members or subgroups within racial group R, and some other meaning M' when applied to members or subgroups within racial group R'.

This claim rests on the assumption that the meaning of an expression can shift depending on what sorts of things it is being used to classify. As we will see in Chapters 2, 3, and 4, I argue that there is a sense in which terms like “criminal,” “thug,” “terrorist,” and “immigrant” have different meanings depending on what racialized groups they describe.

This assumption will strike many orthodox practitioners as preposterous, for two reasons. First, the assumption is highly unlikely to hold for large swaths of natural language. In short, the meaning of “dog” doesn’t change simply because I erroneously use it to refer to cats. Consequently, in making the case that race can drive shifts in meaning-standards, the first goal will be to say what makes racialized vocabulary special.

Yet second, even if it is the case that racialized vocabulary is special in this regard, it’s still far from clear that meaning-shifts are characteristic for this sort of discourse. For instance, even if “terrorist” is more often *used* to classify (*Arab*) *Muslim* bodies and their conduct rather than *White, non-Muslim* bodies and their conduct, “terrorist” still has the same meaning across these two uses. If the meaning of “terrorist” is something like *person who unilaterally engages in politically or religiously motivated attacks on civilians*, this meaning proposes identical criteria for individuals of all races. According to this argument, the fact that *users* of the term “terrorist” apply the term differentially as a result of their racial biases is something to be explained by psychology or epistemology, not semantics.

So, my argument for race-specific shifts in meaning-standards depends on the premises that 1) racialized vocabulary is special in some sense, and that 2) shifts in meaning-standards are

genuinely *semantic* phenomena, and not merely psychological or epistemic phenomena. Let me offer a few words in support of each of these premises, with the understanding that they will be more fully supported as the dissertation progresses.

First, racialized vocabulary is special in the following sense: changes in the *use* of this vocabulary engenders changes in the *meaning* of this vocabulary, and both sorts of changes can engender changes in the natures of the *objects* that this vocabulary describes. Part of the goal of Chapter 2 will be to specify more precisely the sense in which racialized vocabulary is special. But this is already enough to see that an ordinary predicate like “dog” is not special in this sense. Even if we as a culture were to change how we use the term “dog,” and even if the extension of “dog” were to encompass, say, phone books instead, *dogs* themselves would be largely unaffected. They would remain dogs, regardless of what we called them. Since dogs are “indifferent kinds” (Hacking 1999), they do not modify their behavior or identity in response to what we call them. Contrast this with “terrorist.” What sorts of individuals and agents fall into the extension of “terrorist” is highly sensitive to our practices of classification. Yet even stronger, the uses and meanings of “terrorist” affect social reality in a way the uses and meanings of “dog” do not. Were we to change our common understanding of “terrorist,” the terrorists themselves would change, too. Assuming the meaning change was extreme enough, terrorists before the meaning change would truly cease to be terrorists after it.

Second, racialized vocabulary is special because it is *perspectival*. That is, perspectives play a genuinely semantic role in determining the content of this vocabulary. Now, the idea that a class of perspectives could be *metaphysically* relevant by instituting or constructing a property, object, practice, or kind *F* just by regarding, treating, or believing that things are *F-ish*, is a familiar point in the social

constructionist literature (for general accounts see Searle 1995, Hacking 1999, Haslanger 2012, for race-specific accounts see Mills 1998, Haslanger 2000, Taylor 2004, see Glasgow 2009 for discussion.) Yet as we will see in Chapter 2, the possibility that these perspectives might be *semantically* relevant is frequently overlooked, downplayed, or outright dismissed in many mainstream semantic theories. This is unfortunate, since the semantic relevance of perspectives can be used to solve a number of problems. In Chapters 2 and 4 I will show that thinking of racialized vocabulary as sensitive to the deliverances of dominant social perspectives gives us a promising account of generics about racial kinds, as well as a promising account of how racialized terms work.

Race-specific features of the context of use can determine i) the extension of a predicate or the truth-value of a sentence, or ii) the *semantic value* of a predicate or the *truth-conditions* of a sentence. Let's take each of these in turn. I argue that racialized terms like "thug" and "terrorist" take different extensions at different contexts of use. At one context, "thug" might be interpreted as including all and only *non-white violent criminals*, and at another, might be interpreted as including all and only *violent criminals*. Obviously this shiftiness in extension also affects the truth-values of the sentences in which these sorts of expressions are embedded, since the sentence "violent criminals are rioting" might be true and the sentence "non-White violent criminals are rioting" might be false.

Yet I also claim that the *semantic value* of a predicate or the *truth-conditions* of a sentence might vary from context to context. This is a deeper claim, since these forms of "meaning" are usually taken to be what determine extensions in the first place. Consider the sentence "Jerrod is violent," focusing on the term "violent." On my contextualist analysis, this sentence can be paired with quite a few distinct truth-conditions. For instance, perhaps the sentence expresses the proposition that *Jerrod is violent for a third-grader*, or *Jerrod is violent for a criminal*, or *Jerrod is violent for a lawyer*. In other

words, “violent” can be associated with a variety of thresholds for *violence*, and which of these thresholds is relevant for determining the sentence’s truth-conditions is frequently up for grabs at a context. My prediction is that “violent” is a racialized term at a context if the contextual standard or threshold for violence *incorporates racialized properties* (e.g. Jerrod’s racial presentation) as extension-determining aspects of the meaning of “violent.” Chapters 2, 3, and 4 argue that racialized properties are often treated as “parts of the meaning” of a wide variety of lexical items which I call *racialized terms*.

2.3 Conduct-Standards

Finally, racialized discourse both presupposes and propagates race-specific shifts in *conduct-standards*. Conduct-standards can be understood as the norms and expectations that determine our social judgments of the relative acceptability, appropriateness, culpability, praiseworthiness, and blameworthiness of the conduct of ourselves and others. Obviously what sets of conduct-standards we endorse, either as persons or institutions, are going to be highly reflective of our values and interests. Even if we have different values and interests, however, most of us seem to think that the conduct-standards that we end up endorsing ought to deliver *consistent* judgments across situations. For instance, if Jason was wrong to steal donuts in situation 1, Kendra was also wrong to steal donuts in situation 2, as long as there are no morally relevant differences between Jason’s and Kendra’s situations (e.g. if Kendra, but not Jason, were broke and starving to death.) If we were to say that Jason was wrong to steal because he is a man, while Kendra’s theft can be excused because she is a woman, the gender difference alone should strike us as a morally irrelevant

distinction which shouldn't interfere with our holding Jason and Kendra to the same standards of conduct.

However, we frequently fail to be consistent, and adjust our normative and moral judgments based on intuitively normatively or morally irrelevant phenomena. For instance, while virtually all Americans see some social value in not being racist, and most agree that Blacks and Whites are moral equals, these groups are frequently held to varying conduct-standards.⁵

I intend the notion of a race-specific conduct shift to capture three things. First, it is supposed to capture the notion that race can shift judgments of whether an action is right or wrong, acceptable or unacceptable. Second, it is supposed to capture the notion that race can shift judgments of the extent to which an agent or group is *responsible* (culpable, praiseworthy, blameworthy) for performing an action or having a certain trait. Finally, it is supposed to capture the notion that race can shift judgments of how *negative* a response is warranted by a negative action and how *positive* a response is warranted by a positive action.

Race-related shifts in conduct-standards: A judgment J concerning an action A is reflective of a race-related shift in conduct standards if J is part of a pattern of judgments which posit non-accidental relations between racial membership and conduct. This pattern of judgments yield any one of the following modal facts:

a) A is wrong when perpetrated by members of race R but right (or less wrong) when perpetrated by members of race R'

⁵ Lebron (2013) sees this inability to conform to our stated personal and institutional standards "in the face of race" as the core contradiction that both drives continued racial injustice and which is cause for a sense of shame in the face of that contradiction. Lebron's project presents these sorts of contradictions as ultimately moral problems with solutions that require a moral rehabilitation of our "national character." My project is more descriptive. My goal is simply to point to some of the semantic, psychological, and social mechanisms by which race-specific shifts in conduct-standards are reproduced and reflected in our social practices, with the understanding that *undermining* or *confounding* these mechanisms is a task for social theorists and activists alike.

b) Perpetrator of group R, but not a perpetrator of group R' is deemed especially responsible for A when A is negative, and perpetrator R', but not perpetrator R, is deemed especially responsible for A when A is positive

c) Perpetrator R, but not perpetrator R', is deserving of especially negative responses when A is negative, and Perpetrator R', but not perpetrator R, is deserving of especially positive responses when A is positive.

Note that this formulation is disjunctive, and likely does not capture *all* the ways in which race can cause shifts in observed conduct-standards. For instance, I will argue in Chapters 5 and 6 that one core reflection of race-specific shifts in conduct-standards has to do with the attribution of *traits* and the explanation of behavior, rather than token judgments about the moral valence of actions.⁶ In any case, this is a good place to start in getting a systematic read on the phenomenon.

There is a good deal of evidence that an actor's racial presentation can have strong effects on how their actions are socially interpreted. For instance, Duncan (1976) found that White subjects judged an ambiguous action-- a shove-- as more violent when the agent doing the shoving was Black. Moreover, subjects were most likely to judge this action as "aggressive" or "violent" when the agent doing the shoving was Black, and the agent being shoved was White. In a "conceptual replication" of this experiment, Sagar and Schofield (1980) found that both middle-school aged Black and white subjects were more likely to judge relatively innocuous token actions performed by Blacks as more "mean" or "threatening" than the same actions performed by Whites.

This is consistent with findings that racial presentation can affect judgments of blameworthiness. One widely cited study demonstrated that White NBA referees were much more likely to call fouls on Black basketball players during games, while Black referees were more likely to call fouls on White players (Price and Wolfers 2010). These race-specific differential assessments of

⁶ This is because I assume that judgments of the moral valence of actions are bound up with underlying reasoning about the relevant causes and explanations of those actions. I review some evidence for this claim in Chapter 5.

conduct also seem to occur in high-stakes contexts. In a well-known series of experiments conducted by Joshua Correll and collaborators (2002, 2007), the experimenters found that in the context of a shooter simulation program, both trained police officers and the general population share a “shooter bias.” In particular, participants “[decided] not to shoot an unarmed target more quickly and more often when the target is White.” Similarly, participants were quicker and more likely to “shoot” armed Blacks than armed Whites. As the experimenters conclude, “participants seem to process stereotype-consistent targets (unarmed blacks and armed whites) more easily than counterstereotypic targets (unarmed blacks and armed whites.)”

There is also a good deal of evidence that what *abilities* or *capacities* an individual is presumed to have (and not have) is affected by their racial presentation. Let’s consider two examples. In a famous study, Bertrand and Mullainathan (2004) sent out a series of fictitious resumes in response to job advertisements in two major metropolitan areas, Boston and Chicago. The resumes were coded as “high quality” or “low quality,” and each resume was randomly assigned either a stereotypical White name (e.g. “Emily Smith”) or a stereotypical Black/African-American name (e.g. “Lakisha Jackson.”) The experimenters created a four-by-four distribution along two axes: resume quality and name type. This yielded high and low quality resumes with “Black” names, and high and low quality resumes with “White” names. The study found that an arbitrary “White” applicant (regardless of resume quality) was 50% more likely to receive a callback than an arbitrary “Black” applicant. Moreover, while a high quality “White” resume proved substantially more effective than a low quality “White” resume, a high quality “Black” resume was only marginally preferable to a low quality “Black” resume. Intuitively, the standard that an applicant had to meet in order to be considered worthy of a callback differed in decidedly race-specific ways.

Finally, in another reminder that shiftiness in conduct standards is not confined to human interpreters, Angwin et al (2016) report that criminal “risk assessments,” which are sentencing algorithms commonly used by judges in criminal trials to hand down appropriate sentences, deliver assessments which systematically assess Blacks as more likely to commit future crimes than Whites. In an analysis that controlled for gender, recidivism, criminal history, and type of crime committed, the authors found the algorithm treated Blacks as 77% more likely than Whites to commit a future violent crime, and 45% more likely than Whites to commit a future crime of any sort. Since risk assessments are widely used by state and local courts in determining appropriate sentences, this is a serious problem.⁷

3. Racialized Discourse and Social Perspectives

Standard-shifting is the core mechanism at the heart of racialized discourse, and is the key to providing a unified account of its semantic and social-metaphysical manifestations. Understanding in detail how it works in both of these domains may give us some part of a more complete story of how linguistic and social practices reproduce racial domination. In particular, we want to know *why* these various standard shift across the semantic and social-metaphysical domains. The challenge here is to provide a relatively parsimonious theory of the underlying mechanisms that drive these shifts in use-standards, meaning-standards, and conduct-standards which is nevertheless sufficient to do explanatory work at a somewhat fine-grained level. My hypothesis is that standard-shifting always

⁷ As Northpointe Software, the creator of this particular sentencing algorithm, pointed out, this analysis does not control for what one might think of as some of the most important factors that drive differential risk assessments--educational attainment, employment history, and income level, for instance. This is an important point to remember, but it is largely besides the one I am making here. My point is that within the realm of racialized discourse, conduct-standards shift according to racial presentation. Given that an individual’s “racial presentation” is seldom isolated from facts about that individual’s socioeconomic status, educational opportunities, geographical location, and the like, we should also expect these sorts of factors to play a role in how the conduct-standards are shifted.

takes place from the point of view of a *social perspective*. Consequently, my second core claim is that *shifts in use-standards, meaning-standards, and conduct-standards are driven by the presence of dominant social perspectives which give special weight to racialized distinctions.*

In *Black Power*, Charles Hamilton and Kwame Ture (b. Stokely Carmichael) provide a useful series of examples illustrating how social perspectives might play a role in shifting these standards. In the midst of a discussion of how the dominant “white society” shapes historical memory, they note:

[Political control] includes the attempt by the oppressor to have his definitions, his historical descriptions, accepted by the oppressed... The recorded history of this country’s dealings with red and Black men offers other examples. In the wars between the white settlers and the “Indians,” a battle won by the Cavalry was described as a “victory.” The “Indians” triumphs, however, were “massacres.” (The American colonists were not unaware of the need to define their acts in their own terms. They labeled their fight against England a “revolution”; the English attempted to demean it by calling it “insubordination” or “riotous.”)... Black people came to be depicted as “lazy,” “apathetic,” “dumb,” “shiftless,” “good-timers.” Just as red men had to be recorded as “savages” in order to justify the white man’s theft of their land, so Black men had to be vilified in order to justify their continued oppression. (Hamilton and Ture 2011: 35-36)

This passage nicely captures the ways in which dominant social perspectives can shift use-standards, meaning-standards, and conduct-standards. First, the authors recognize that staking out and controlling key pieces of vocabulary not only institutes norms for the use of that vocabulary, but also institutes perspectival norms for interpreting the *conduct* of the groups being described by that vocabulary. Creating this discourse on the “Indians,” for example, is a vital tool not only in getting “[the oppressor’s] historical descriptions accepted by the oppressed,” but also in getting previously uncommitted third parties to adopt a perspective which includes the “fact” that “Indians” are *savages* who engage in *massacres* of “Americans.”

Second, the recognize that the meanings of many predicates-- including “massacre,” “lazy,” and “revolution”-- shift along with the nature of the perspective being taken up. When a social

perspective becomes dominant-- for instance, one which sees Blacks as lazy and shiftless-- then lexical items like “lazy” and “shiftless” come to be infused with racialized meanings (in a sense to be made clear.) Since perspectives are partially meaning-determining, which perspective wins out is highly consequential for what such terms and concepts which *mean* in a variety of contexts. Hamilton and Ture endorse a rather relativistic interpretation of this idea when they quote Lewis Carroll’s *Through the Looking Glass* just a few lines later⁸:

“When I use a word,” Humpty Dumpty said in a rather scornful tone, “it means just what I choose it to mean--neither more, nor less.” “The question is,” said Alice, “whether you *can* make words mean so many different things.” “The question is,” said Humpty Dumpty, “which is to be master-- that’s all.”

Finally, the authors note that which descriptive vocabulary it is considered legitimate to apply to a group is partially determined by perceptions of what rights and entitlements that group has. For instance, coordinated political action by a population counts as a “revolution” only insofar as that population is recognized to have a *right* to cast off an oppressive government. The very same political action might be described as “riotous” or “insubordinate” insofar as that population is not perceived to have such a right. In short, this is an instance of conduct-standards informing meaning-standards and use-standards.

Ture and Hamilton’s insight that perspectival positioning can shape how one sees political realities provides a promising way of thinking about race-specific shifts in use-standards, meaning-standards, and conduct-standards work. I will argue that social perspectives play two distinct roles, each of which corresponds to each domain of racialized discourse.

⁸ In Chapter 2, I make clear that this sort of semantic perspective-dependence is not a global feature of natural language, but is characteristic of a circumscribed vocabulary that I call “socially perspectival predicates.”

Recall that in the semantic domain, we are trying to explain how linguistic use-standards and meaning-standards shift as a result of taking on a certain social perspective. So first, I will argue that social perspectives play a role in *the online interpretation of racialized discourse* (and perhaps other forms of discourse as well.) The basic idea is that when an interpreter is computing the content of a linguistic string, they occasionally must consult a *privileged* perspective in determining features of the meaning of that string. For instance, an interpreter needs to know something about the geopolitical perspective in play in determining what proposition the following sentence expresses:

1) Tina is a foreigner.

In assigning a semantic value to 1), the problem that the interpreter has to solve is this: which nation is Tina foreign *to*? The answer to that question changes depending on which geopolitical perspective is treated as most relevant at the context, and key elements of what is being said are left out if the interpreter does not have a grasp on which of these perspectives are being honored.

In Chapters 2, 3, and 4 I argue that this same sort of perspectivalism applies to examples of racialized discourse.

In the social-metaphysical domain, we are trying to understand how dominant social perspectives play a role in fixing differential social expectations and different norms of conduct for different sorts of racialized bodies. At this level, social perspectives are invoked to explain how members of social structures converge on *dominant understandings* of the nature and conduct of racialized categories. Social perspectives thus play two rather different roles: in the semantic domain, our account of social perspectives ought to explain how those perspectives can play an empirically tractable role in properly *linguistic* interpretation. Competent language users navigate these

perspectives “on the fly,” as it were, in everyday linguistic exchanges. Yet in the social-metaphysical domain, we interested in the role that social perspectives might play in widespread “extra-linguistic” interpretations of racial reality. To put things simply in the terms of our opening anecdotes, we ought to distinguish the semantic question--what role do social perspectives play in explaining “thug”’s potential to mean “Black” at some level relevant to semantics and pragmatics?-- from the social-metaphysical question-- what role do social perspectives play in forging a non-accidental link between Black conduct and (what counts as) thuggish conduct in a social structure?

4. Structure of the Dissertation

The dissertation is divided into 7 chapters. Chapters 2-4 develop a semantic framework which gives content to the idea of race-specific shifts in meaning standards. Chapter 5 explores how these race-specific shifts are reflected in thought about racialized social categories. In particular, it shows how racialized categories are *essentialized* in common sense conception, and how essentialist patterns of reasoning function to uphold social hierarchies. Chapter 6 shows how the pragmatic profile of racialized discourse allows it to distribute these essentialist representations in everyday linguistic exchanges. Finally, Chapter 7 argues that racialized discourse can create new social categories, and in so doing, perpetuate racial domination.

Chapter 2: Generics, Race, and Social Perspectives

The goal of Chapter 2 is to motivate the claim that some linguistic items are “socially perspectival” in the sense outlined above, with the goal of showing that many instances of racialized discourse are also socially perspectival. Rather than motivating such a framework in the abstract, we

begin with a concrete test case: *generics*. Generics are kind-wide generalizations such as “Ravens are Black,” “Birds fly,” and “Moss grows on the north side of trees.” After a discussion of some of the basic features that generics have in common with one another, I propose to investigate a class which I call “Type C generics.”

Examples of this class include sentences like “Blacks are criminal” and “Muslims are terrorists.” I argue that the meanings of Type C generics are sensitive to race-specific meaning shifts-- these sentences are interpreted differently depending on what racial groups they are about. My claim is that when embedded in certain generic constructions, lexical items like “criminal” and “terrorist” receive default racialized interpretations in which these terms conventionally token racial or ethnic identities. Such interpretations can be explained in terms of an interaction effect between the formal properties of generic language on one hand, and the nature of racialized terms (and perspectival discourse) on the other.

This explanation contrasts with another recent explanation of the meaning and interpretation of social kind generics. Sarah-Jane Leslie (2017) and her collaborators (Khemlani et al 2009, Rhodes et al 2012) have demonstrated that many “striking property” generics are treated as true even when very few members of the category instantiate the striking property in question. For instance, most interpreters accept “Mosquitoes carry the West Nile virus” as true, even though less than one percent of all mosquitoes carry the virus. Other examples include “pitbulls maul children” and “sharks attack swimmers.” Within Leslie’s (2007, 2008) semantic framework, “prejudicial” generics such as “Muslims are terrorists” and “Blacks are criminal” are basically treated on the same model as “striking property generics.”

I believe that there is something fundamentally wrong with Leslie's conflation of striking property generics and what I will call "Type C" generics, but to show this requires some work. Leslie's semantic framework is driven by the widespread assumption that generics are fundamentally generalizations which describe the non-linguistic world as it is "independently" of that generalization, and they are true to the extent that they adequately represent that world. As I will argue, the problematic part of the assumption is this: the *semantic* status of the generic-- the question of whether it is true or false-- is frequently treated as isolated from *cultural* or "worldly" (Leslie 2007) facts about what the generic is broadly taken to mean within a discourse community, or what the generic is typically used to say within that community.

When it comes to racialized discourse, we cannot afford to isolate a theory of the meaning of linguistic symbols from a story about what those symbols are leveraged to classify. In making this argument, I first identify a particular subclass of terms which I dub *socially perspectival predicates* (SPPs.) SPPs have three features in common. They 1) pick out *perspective-dependent* objects and kinds, 2) require *conventional* epistemic standards to determine whether a candidate object fits into the extension of an SPP, and 3) have patterned uses which are capable of *instituting* the objects and kinds which SPPs pick out. I argue that in computing the content of an SPP, interpreters must make implicit reference to socially privileged perspectives which are treated as authoritative about whether a given object fits into the extension of the predicate. Such deference grants these authorities influence over both the conventional meaning of these terms and over the nature of the objects in the social ontology that these terms purport to describe, much the way a baseball umpire is authoritative over the meaning and metaphysics of "strike"/*strike*.

“Racialized terms” are socially perspectival in this sense. I suggest that the interpretation and use of Type C generics occurs against the backdrop of socially transmitted beliefs about racialized groups, and that the online interpretation process is penetrated by these background beliefs. Racialized SPPs like “terrorist” provide an instruction to the interpretive system to consult a social perspective which determines the extension these terms take at a given context. Certain perspectives which have emerged as socially dominant-- for instance, a perspective which “sees” Muslim identity and terrorism to be significantly overlapping classes-- are most likely to be consulted during the online interpretation process.

Chapter 3: When Code Words Aren't Coded

Chapter 3 places what I am calling “racialized terms” front and center. Examples include “terrorist,” “immigrant,” and “thug,” whose definitional senses make no reference to race, but which are frequently used to make claims about *(Arab) Muslims, Latinos, and Black/non-White men* under a thin veneer of descriptive race-neutrality. Since the use and interpretation of racialized terms in everyday discourse is frequently both product of and causal contributor to harmful forms of racial ideology, there are strong socio-political as well as semantic motivations to get clear on what racialized terms mean, and how they manage to convey their characteristic messages.

I argue that racialized terms display systematic race-related shifts in meaning-standards. In particular, the *referential content* of these terms shifts depending on which racial groups are being described with these terms. I argue that the observed racial norms within specific discourse communities can constrain the meanings of racialized terms. These norms are thus one source of what I have called race-related shifts in meaning-standards.

In defending this claim, I intervene into a contemporary debate about racial communication in political speech. The problem is this: politicians used to have to rely on “code words” to get their racial messages across. Yet recent years have demonstrated that politicians often get away with quite explicit forms of racism in mainstream political discourse. It doesn’t seem that much of this speech is “coded” anymore. On the other hand, Americans in general do seem to think that racism is bad, and are hesitant to endorse attitudes that they believe to be racist. How can we reconcile the effectiveness of explicitly racialized speech with Americans’ reticence to be racist?

I address this question in two steps. Negatively, I argue that the “standard framing” of racially coded speech, presupposed by much extant work in philosophy of language, is insufficient to account for the normative structure of more recent (and more explicit) racial appeals. In essence, the standard framing fails to appreciate the fact that many interpreters are incentivized to defect from more “standard” anti-racist norms in favor of the racial norms extant in their own discourse communities. More positively, the chapter diagnoses some of the mechanisms that underlie such defection. I provide an analysis of an ascendant type of explicitly racial appeal called candid racial communication. I proceed to argue that we should see the normative structure of candid racial communication as having important upshots for a theory of what racialized terms mean.

Along the way, we will discuss the political function of racialized terms. I argue that in strategically using racialized terms, speakers invite their audiences to take up a perspective which is reflective of *that audience’s* self-conceptions and social interests as they perceive them. Getting an audience to believe or invest in what is being communicated via racialized terms is largely a matter of getting those audiences to see what they are or ought to be committed to as a function of how they perceive *themselves*. To these audiences, the question of whether they are violating any norms

prohibiting racial bigotry in responding to not-so-coded racial messages is far less important than whether these responses include them within the group that is being appealed to.

Chapter 4: The (Meta-)Semantics of Racialized Terms

Chapter 3 is concerned with specifying how racial norms shape the *referential content* of racialized terms. Chapter 4 stakes a much stronger claim: racialized terms also display race-specific shifts in meaning-standards at the level of their *standing meanings*. As we will see, this conclusion is entailed by plausible metasemantic assumptions.

An account of the meaning of racialized terms should do two things. First, it must explain why racialized terms propose different *thresholds* for different sorts of racialized agents. That is, the “semantic boundaries” of racialized terms expand and constrict depending on what sorts of racialized individuals and groups are under discussion. Second, the racialized “content” of these terms is *defeasible*. It is possible to use a racialized term at a context while disavowing commitment to any particular racialized content.

In response to the threshold problem, I argue that to apply a racialized term to an individual is to locate that individual above a certain threshold corresponding to the term. In such predications, the racialized properties of the individual or group being described play a role in determining whether or not that individual indeed is above that semantic threshold. In essence, this predicts that whether or not an individual is deemed to be in the extension of “thug” frequently has a lot to do with what that individual’s racialized properties are.

In response to the defeasibility problem, I contend that we can diagnose a number of explanations for the defeasibility of racialized meaning if we embrace three metasemantic

assumptions: racialized terms are i) socially perspectival, ii) context-sensitive, and iii) meaning-underdetermined. I flesh out these assumptions and argue for their plausibility. I then show that the relatively “plastic” conception of linguistic meaning that my framework develops need not threaten the project of delivering an informative and systematic first-order semantics for racialized terms. In demonstrating this, I develop my own first-order semantics. I show that many racialized terms can have *inherently generic* logical forms, and that this hypothesis allows us to identify manipulable elements in the syntax of racialized terms whose semantic values can be fixed by context. Taken jointly, Chapters 3 and 4 show that racialized terms are subject to race-related shifts in meaning-standards at the level of “sense” as well as the level of “reference.”

Chapter 5: Essentialism, Racialized Categories, and Moralistic Attribution

Chapters 2-4 focus squarely on how the meaningful features of *linguistic* representations can shift systematically in the face of race, and how those representations can be used in ways in communicate substantive racialized messages. Chapter 5 focuses on the question of how the meaningful features of *mental* representations can shift in the face of race.

Drawing on empirical evidence from social and developmental psychology, I propose that many social categories-- including *criminal*, *terrorist*, *immigrant*, and *welfare recipient*-- are commonly racialized in common sense conception. When they are racialized, they are particularly liable to being *essentialized*: they are implicitly understood to instantiate a non-obvious essential nature which is causally responsible for category members’ socially pathological behaviors.

I argue that social categories that are *both* racialized and essentialized are particularly prone to what I will call *moralistic attribution*. The idea is that when the category of *welfare recipient* is perceived as

“Black” or “non-White,” the negative behaviors of welfare recipients are more likely to be seen as attributable to dispositional moral failings. Conversely, positive behaviors of these recipients are more likely to be seen as attributable to situational, non-dispositional factors. When categories are *not* racialized in specific ways, moralistic attribution is less likely.

At first glance, the notion that a category is essentialized seems to be in tension with the notion that its behaviors are moralized. After all, if an agent cannot help but act in accordance with their underlying essence, on what grounds can they be held responsible for their behavior? I show that in the case of racialized social categories, this tension is only apparent. As it turns out, essentialists about social categories are happy to treat both natural, immutable *and* non-natural, mutable mechanisms as the “essence” in question that defines the social category. For instance, anti-Black racists perceive “Black criminal pathology” to be rooted *either* in an immutable natural essence or in an essentialized “defective culture.” Essentialists about (Black) criminality will thus endorse either explanation when they want to make moral judgments about (Black) criminals.

I intend the moralistic attribution error (MAE) to serve as a working hypothesis. Yet formulating that hypothesis gives us a chance to explore interesting points of convergence and overlap between at least three distinct projects in social psychology: i) the project of tracking the *explanatory strategies* individuals rely on in making sense of others’ behavior, ii) the project of accounting for the genesis of the *responsibility attributions* that guide individuals’ moral judgments, and iii) the project of understanding the complications that *intergroup cognition* introduces to i) and ii).

Chapter 6: Racialized Discourse and the Flow of Information

The goal of Chapter 6 is to make clear how racialized discourse can trigger the sorts of representations and prompt the sorts of attributions discussed in Chapter 5. In so doing, we will i) model the characteristic effects that generics and racialized terms have on the flow of information among interlocutors and ii) describe the sorts of inferences and conclusions that such interlocutors draw on the basis of racialized linguistic inputs. Following the semantics for racialized terms in Chapter 4, I argue that we can provide a unified account of how both Type C generics and racialized terms affect the common ground if we treat the latter as “inherent generics.” The production and interpretation of both Type C generics and racialized terms can then be seen to have cognitive effects which threaten to decrease social regard for the agents described via this sort of language.

The account of how racialized discourse triggers socially marginalizing attitudes falls out of my account of how racialized discourse affects the flow of information among speakers and hearers. Following the communication model pioneered by Robert Stalnaker’s conception of “common ground,” I argue that generic language can impose orderings on possible worlds, understood as conversational “live options” within the context set. These ordering mechanisms allow racialized generic language to serve what I will call *essentializing-alethic* and *legitimizing-preferential* conversational functions--it carries information about how the world *is* and about how it would be good for the world to be.

In supporting this seemingly counterintuitive hypothesis, we draw on the resources of system justification theory to illustrate the ways in which interpreters take what is *normal* to be (defeasibly) *preferable* to “abnormal” alternatives. As such, worlds in which *Blacks are criminal* or *Muslims are terrorists* are regarded not only as more *likely* to be true, but also worlds which are *more normal than* (and therefore preferable to) worlds in which these states of affairs do not hold.

Chapter 7: Constructing Racialized Categories

The argument of Chapters 2-6 represents an attempt to model how racialized discourse transmits information and how it functions in communicative and cognitive practices. What remains to be demonstrated is how these communicative and cognitive practices might play a role in sustaining racial stratification.

Chapter 7 closes the dissertation by articulating how racialized discourse plays a role in creating social role positions that become racialized within the context of concrete social practices. I introduce the notion of a *shifty racialized category* (SRCs.) SRCs are the categories picked out by the terms we have been discussing all along: “thug,” “terrorist,” “immigrant,” and the like. Importantly, to say that a category is racialized is not to say that being a member of a particular race is necessary and/or sufficient for category membership. Consequently, I show that certain social categories can “track” racial features without it being the case that every member of the category is racialized in the same way.

The tracking relation denotes a relationship between category membership and certain racialized properties distributed across a population. Drawing on a historical example of the progressive racialization of the social category *criminal*, I show that facts about racial category membership often non-accidentally co-occur with facts about membership in the category *criminal*. This co-occurrence calls for an explanation. SRCs provide partial explanations of the principled *co-occurrence* of facts about category membership and facts about racial presentation. Here’s how: when a non-racialized category C is racialized in “ambiguous” ways, an SRC is instituted. During this

institution, the socially observed conditions for membership in the original C are “split” between the original grounding conditions and new, race-specific grounding conditions.

Social categories are SRCs to the extent that it is often underdetermined which grounding conditions are in play. One way of getting a read on these grounding conditions is provided by Ásta Sveinsdottir’s “conferralist” account of how social categories are constructed by our linguistic practices. After demonstrating the utility of Ásta’s framework for getting clear on the structure of SRCs, I close by showing that that framework gives us some normative hints about how we might be able to challenge the linguistic bases of racial domination.

Chapter 2. Generics, Race, and Social Perspectives

Abstract

In this chapter, I argue that a wide variety of generics about racial kinds are semantically context-sensitive. I argue that generics such as “Muslims are terrorists,” “Blacks are criminal,” and “Immigrants are violent” are context-sensitive because the lexical items “terrorist,” “criminal,” and “violent” take different extensions at different contexts. Call these lexical items *racialized terms*. One of the central contextual features that allows them to take these different extensions, I argue, is a race-related shift in meaning-standards. The source of these race-related shifts are what I call *dominant social perspectives*. In online interpretation, naive interpreters adjust their interpretations of these terms in accordance with the deliverances of dominant social perspectives that affect the context of use.

This positive view develops through an engagement with Sarah-Jane Leslie’s influential “weak semantics” for generics. I show that Leslie’s semantics falls short on two counts. First, it does not appreciate the extent to which the meanings of racialized terms are determined by their role in social practices. Second, it does not adequately explain the phenomenon of race-related meaning shifts; in this case, why generics such as “Blacks are criminal” and “Immigrants are rapists” are more readily accepted than generics like “Whites are criminal” and “lawyers are rapists.” Finally, I propose a semantic framework meets both these desiderata, and extend it in Chapter 3. As we will see, this framework is rather unorthodox from the perspective of most mainstream semantic traditions. In particular, it entails treating many phenomena often believed to belong to the domain of epistemology or metaphysics as proper objects for *semantic* theorizing.

1. What are Generics?

Generics are generalizations about kinds such as:

- 1) Tigers are striped.
- 2) Birds fly.
- 3) Mosquitoes carry the West Nile virus.

Despite their ostensible innocence, the problem of determining the truth conditions of generics has bedeviled semanticists for over 40 years. Part of the problem is that generics are not susceptible to any uniform quantificational analysis. For instance, despite the fact that 1)-3) are each intuitively true and have the same surface structure, each sentence is true under different quantificational conditions. Sentence 1) seems to express something like *all tigers are striped* and thus requires a universal quantifier. On the other hand, 3) seems to be existentially quantified--it is true if there are at least some mosquitoes which carry West Nile. Finally, sentence 2) seems in need of a quantifier like "most." After all, most birds do fly, but penguins, ostriches, and over 40 other species of bird do not.

In fact, generic statements do not behave like kind-wide quantifications at all. For one thing, even paradigmatic "majority generics" like 1) are not equivalent to any universally quantified sentence:

- 1) Tigers are striped.
- 1a.) All tigers are striped.

Notably, 1), but not 1a.), is compatible with the existence of non-striped tigers. Other, more tricky generics abound:

- 4) Ducks lay eggs.

5) Ducks are female.

6) Books are paperbacks.

Intuitively, 4) is true, and 5) is false. Why? After all, there are more female ducks than there are egg-laying ducks, since roughly half of ducks are female and a subset of that are unable or too immature to reproduce. 6) is intuitively false, despite the fact that 80% of all books are paperbacks. Thus the sheer *prevalence* of a property F among a kind K does not seem to be evidence for the truth of a generic of the form *Ks are F*.

Most theorists assume that all generic sentences share a common property called *genericity*. On this approach, studying generics is a matter of finding out what genericity is, and then letting the metaphysical account of genericity determine the semantics for generics. I will have little to say about genericity, since I am not optimistic that genericity is a unified phenomenon, and thus not optimistic that a single semantic framework adequate to model the truth-conditions of *all* generics is possible.⁹ Consequently I will not be exploring the semantics of generics *tout court*, but will confine myself to a semantic account of a special class of generics that I will dub "Type C Generics."¹⁰ With

⁹ This is not to say that ingenious and heroic global analyses of genericity and generic truth-conditions have not been proposed. The three dominant approaches have been *normality* and/or *possible worlds* approaches (Asher and Moreau 1995, Nickel 2016), *stereotype* theories (Geurts 1985, Declerck 1986), and *probabilistic* approaches (Cohen 1999). Roughly speaking, normality approaches attempt to cash out the truth-conditions of generics in terms of some notion of normalcy. For instance, a generic statement "As are F" might be interpreted as saying that *normal* As have property F, or that As do or would have property F in some set of *normal* possible worlds. Stereotype theories hold that "As are F" is true iff it is the case that the *stereotypical* A has property F. Probabilistic approaches associate the generic truth-conditions of "As are F" with "comparative probabilities" which specify 1) the probability of an arbitrary *x* in A having property F or 2) the probability of an arbitrary *x* in A having property F as opposed to the probability of an arbitrary *y* in a superset of A having property F. Either strategy is employed depending on what the intuitive truth conditions of the generic statement are. Finally, see Sterken (2015b) for an *eliminativist* approach to genericity: while there is no such thing as genericity, generics have truth-conditions nonetheless.

¹⁰ As will become obvious presently, the moniker "Type C" generics is a nod to Sarah-Jane Leslie's own taxonomy consisting of "Type A" and "Type B" generics.

a few notable exceptions (McConnell-Ginet 2003, 2012 Haslanger 2011, Leslie 2017), the generics literature has largely avoided this interesting and problematic class.¹¹ Examples include:

- 7) Blacks are criminal.
- 8) Immigrants are violent.
- 9) Muslims are terrorists.

I will be concerned to show that Type C generics have two important features. First, Type C generics link social kinds with special lexical items which I will call *socially perspectival predicates* (SPPs).¹² Roughly, SPPs are special because they require interpreters to make reference to socially privileged perspectives which are authoritative about whether a given object fits into the extension of a given SPP.

Second, sentences like 7)-9) trigger specific interpretations of the predicates attributed to them. These interpretations of SPPs like "criminal," "thuggish," and "terrorist" are *racialized*-- they link criminality, thuggishness, and terrorist activity to specific racial and ethnic identities, and imply that this link is nomic or non-accidental. For instance, a racialized interpretation takes whatever properties are denoted by the term "terrorist" (e.g. being a fanatic, being foreign, being violent) to be non-accidentally and characteristically instantiated by *Muslims*, as opposed to members of other groups.

¹¹ This omission is not surprising among authors primarily concerned with the truth-conditions of generics, since these authors take sentences like 7)-9) to be uniformly false, or as entailments of severely flawed theories about the world (e.g. Nickel 2016). Consequently, most of the work on these sorts of generalizations has been pursued by those primarily interested in why certain generics are *treated as true* by theoretically unsophisticated language users. This is also my starting point, and it seems to be shared by theorists such as Leslie (2007, 2008) and Haslanger (2011, 2014). We begin with the fact that these generics are widely accepted *regardless* of what their "actual" truth value is, and assemble a semantic/pragmatic framework which explains why these generalizations can *seem true* to those interpreters.

¹² The "social kind" condition indicates that many Type C generics include non-racialized kinds. Examples include "Women are submissive," "Men are pigs," "Police officers are sexist," etc. I won't make the argument here, but I believe that my semantics can be extended to cover these cases.

Importantly, I will not be interested here in the question of whether Type C generics are true or false *simpliciter*. Rather, we will begin with the facts that i) Type C generics are widely accepted, and ii) that the SPPs within Type C generics receive racialized interpretations. That is, the data that we are trying to account for is that Type C generics are widely *treated as true*, not that they *are* true. In what follows, I argue that the best way of accounting for these facts is that i) Type C generics are context-sensitive, and ii) that the meanings of terms like “thug,” “terrorist,” “criminal,” and the like are sensitive to dominant social perspectives. In making this argument, we'll turn to Sarah-Jane Leslie's influential “weak semantics” for generics, which provides both a unified theory of what genericity is, and a global account of the truth-conditions of generics. I argue that despite the virtues of her account, Leslie's semantics cannot give an adequate explanation for the behaviors of Type C generics. In the first place, her account fails to draw relevant distinctions between the semantic profile of generics which describe states of affairs independent of human social practices, and generics which describe states of affairs which are partially constituted by human social practices. And in the second place, Leslie's account does not have the conceptual resources to explain how Type C generics prompt racialized interpretations.

2. Leslie's Weak Semantics: Type A and Type B Generics

2.1 Genericity and primitive generalizations

In her groundbreaking work on generics, Sarah-Jane Leslie has argued that the capacity to understand and interpret generics is a byproduct of a primitive generalization mechanism which is involved in early conceptual categorization and inductive reasoning. In fact, this is just what genericity is, on Leslie's view. All generic sentences share the fundamental property of being

expressions of the primitive generalization mechanism. Consequently, a semantics for generics ought to begin from this basic fact, and must be constrained by facts about how interpreters-- both children and adult-- form generalizations about the world.

If Leslie is indeed correct that generics constitute indirect evidence for the existence and function of a primitive generalization mechanism, we ought to expect children's competence with generics to manifest itself fairly early in cognitive development. A variety of empirical data lend *prima facie* support to Leslie's hypothesis. As early as 3 years old, children display adult-like competence with recognizing, interpreting, and reasoning with generic language (Gelman 2003). For instance, they seem to be sensitive to the ways in which the truth-conditions of generics depend on the precise nature of the property being attributed to the kind. Preschoolers will accept claims such as "lions have manes" while rejecting claims such as "lions are boys," "despite implicitly understanding that there are at least as many "boy" lions as there are maned lions." (Brandone et al 2012, Leslie 2014.)

On the other hand, preschoolers are far less competent when it comes to sentences involving quantifiers. Full competence with quantifying words such as "all," "most," and "some" arises comparatively late in development (Gelman 2003, Halberda and Feigenson 2008.) If we assume that computing more complex truth conditions involves more complex computations, this is a fairly surprising result. After all, the semantic profile of quantifiers seems to be far simpler and more uniform than the semantic profile of generics. Intuitively, "All ravens are black" is satisfied just in case every raven in a given domain is black, "most" is satisfied just in case more than half of all ravens are black, "some" is satisfied just in case at least one raven is black, etc, whereas most semantic theories hold that the truth conditions of generics are rather complex. Surely we should

expect comprehension of generics to piggyback on comprehension of quantifiers! Not only is this not the case, but it turns out that both children and adults *recall* quantified sentences as generics, suggesting that generics are both more cognitively fundamental and more easily interpreted than quantified sentences (Leslie and Gelman 2012.)

2.2 *A Weak Semantics for Generics*

Despite the unruly semantic behavior of generics, adults and children seem to be in possession of a procedure for computing their truth-conditions, even when it comes to aforementioned “tricky” generics like “ducks lay eggs” or “mosquitoes carry West Nile.” Consequently, Leslie holds that the semantic profile of generics out to be i) rather simple and uniform, yet ii) flexible enough to reflect that the intuitive truth-conditions of generics are highly sensitive to the nature and scope of the kinds and properties being generalized about.

Leslie divides “tricky” generics into two types, Type A and Type B. Type A generics are intuitively true in virtue of a) predicating properties which belong to a certain *characteristic dimension* for the kind and b) predicating properties whose counter-instances are *negative* rather than positive. Plausible candidates for “characteristic dimensions” include body morphology, reproduction, vocalization (in the case of animal kinds), as well as functional roles and perhaps cultural norms in the case of conventional or social kinds. The distinction between positive and negative counter-instances ultimately amounts to a difference in psychological salience:

A positive counterinstance to Ks are F occurs when an instance of the kind K has a concrete alternative property, that is, when it has a *positive alternative* to the property F , while negative counter-instances occur when an instance simply fails to be F . Whether a counterinstance counts as positive or negative is highly dependent on the property being predicated. (Leslie 2007: 66)

If a generic predicates a property with salient *positive* counter-instances, then the generic is likely to be judged false. If the counter-instances to the property predicated are negative, then the generic is likely to be judged true. For instance, there is a positive alternative to being *female*; one could be *male*. On the other hand, it is not clear that there is any positive alternative to *laying eggs*, at least for ducks. Thus, "ducks lay eggs" is judged true because laying eggs lies along a characteristic reproduction dimension *for* ducks, and there are no positive counterinstances. Conversely, "ducks are female" is judged false because there is a positive alternative property to being female.¹³

Generics of Type B, on the other hand, are intuitively true in virtue of ascribing "striking properties" to kinds. Striking properties are harmful, frightening, or dangerous properties which we would be wise to avoid. Examples include "Mosquitoes carry West Nile" and "sharks attack swimmers." These generics are frequently judged to be true even when the objective prevalence of the property among the kind is very low, since the cognitive cost of treating these generics as true is far lower than the possible risk of treating them as false. Consequently, people tend to overestimate the prevalence of a striking property among a relevant kind, and these sentences tend to receive a weighted role in inference (Khemlani et al 2009). Moreover, the knee-jerk overgeneralizations which Type B generics prompt seem to play a role in the formation of prejudices and stereotypes about social kinds (Leslie 2017, Rhodes et al 2012).

Leslie takes the behavior of Type A and Type B generics to be strong evidence for two claims. First, our facility with these types show that generics express "cognitively fundamental generalizations" about kinds. The seemingly disordered character of our semantic intuitions about

¹³ However, see Sterken (2015a) for counterexamples to and arguments against the claim that all true Type A generics have negative counter-instances.

tricky generics show that this fundamental generalization mechanism is rife with biases which affect the space of kinds and properties over which we generalize. Second, the semantics of the GEN¹⁴ operator must be relatively simple, since even young children can grasp the semantic contribution made by the generic operator, even though it can be attached to all sorts of different sentences. A uniform metalanguage interpretation of GEN thus cannot account for the truth-conditional diversity that Type A and Type B generics display. Accordingly, Leslie opts to simply disquote GEN in the semantics, offering no interpretation of GEN in the metalanguage:

“Tigers are striped” is true iff [GEN] Tigers are striped.

Note that like any disquotational theory, this clause black boxes many questions about what the world must be like in order for “tigers are striped” to be a true generic. For instance, the clause says nothing about *how many* tigers need to be striped, *which* sorts of tigers need to be striped, what counts as being *striped*, and so forth. From Leslie’s point of view, as well as from the point of view of

¹⁴ According to the dominant trend in semantics, the logical form of generic sentences is very close to the logical form of sentences which include adverbs of quantification such as "usually," "generally," etc. (Lewis 1975, Krifka et al 1995, Leslie 2007.) In particular, sentences of this sort have a *tripartite structure* which includes a variable-binding Operator of some sort, a Restrictor, and a Matrix:

Operator x,*.....*,z [Restrictor x*.....*z] [Matrix x*.....*z]

The Restrictor specifies the objects to which bound variables are assigned. The Matrix specifies the property or properties which are assigned to the objects in the Restrictor. The job of the Operator is to bind the variables in the Restrictor and Matrix and to relate them to one another. In the case of sentences like "All ravens are black" or "Some students are responsible," the Operator will be a universal or existential quantifier. In the case of generics, the Operator will be a variable-binding Operator GEN. Bracketing the question of what the semantic contribution of GEN is, the logical form of "Tigers are striped" will look like this:

1b.) GEN [Tiger(x)] [Striped (x)]

virtually every mainstream approach to generics, these are not questions for a semantics for generics, but *metaphysical* questions for a theory of genericity.

Thus while Leslie's semantics for generics is rather simple, her theory of genericity is rather complex. That theory (Leslie 2007, 2008) specifies the set of conditions which any generic sentence must meet if it is to be judged true:

A generic *K's are F* is true iff:

1. The counter-instances to the claim are *negative*. That is, there is no noteworthy or salient property G such that K's which are not F are G.
2. If F lies along a "characteristic dimension" for K's, then some K's are F. If K is an artifact or social kind, then F is the "function or purpose" of kind K.
3. If F is a striking property, then some K's are F, and all other K's are disposed to be F. If F is not a striking property, [and if "K's are F" is not a minority Type A generic,] then almost all K's are F.

Again, these clauses do not deliver *semantic* truth conditions for generics. Rather, they are "worldly truthmakers" which specify how the world must be in order for the generic to be true.

For Leslie, the distinction between semantic truth and worldly truthmakers is of the utmost importance. If the semantics of generics were as complex as these truthmakers, we would not expect children to display competence with generic interpretation and generic truth value judgments until much later in their development. More importantly, the distinction between the semantic question of what sentence X *means* and the metaphysical question of what the *world* must be like if sentence X is to be true seems to be rather fundamental.

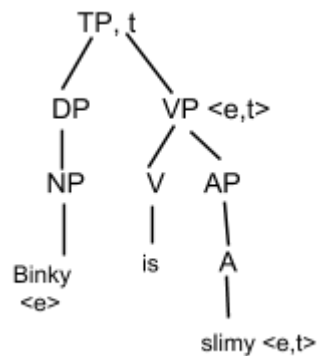
For example, take the example of what it is to be *slimy*. Semantic questions would include: what does "slimy" mean? What kinds of objects are in the extension of "slimy"? Metaphysical questions would include: what facts about an object make it *slimy*? Is *sliminess* a property of the world

external to our senses? Or is it something that only exists if there are creatures which can experience *sliminess*? Now consider the following sentence:

10) Binky is slimy.

In providing a semantics for 10), we can punt on the metaphysical questions of whether *sliminess* is "out there" in the world, or somehow a "product" of our experience, or whatever. All we're interested in is the sentence's compositional structure and its truth conditions. And these semantic features can be provided rather easily, using the standard resources of generative linguistics and first order logic:

Compositional Structure of 10):



Truth conditions 10): Slimy(Binky)

On the other hand, if we are going to specify the conditions which must be in place if 10) is going to be true, we need to choose a specific metaphysical view of *sliminess*. For the sake of argument, let's say that in order for an object to be slimy, that object must be *experienced* as slimy by a normal observer in normal perceptual conditions. Call this "the dispositional theory of sliminess." This theory says nothing about what "slimy" *means*, but only says something about what it is for an

object to actually *be* slimy. Unlike our analysis above, it doesn't give us the truth-conditions of 10), nor its compositional structure or logical form.

What is important for our purposes is that *even if* the dispositional theory of sliminess is correct, that would explain nothing about how we can be competent with the sentence "Binky is slimy." We can very well understand what the sentence says (once we fix a referent for "Binky") without having any metaphysical theory of sliminess. If the dispositional theory of sliminess *did* deliver a semantics for "slimy," the theory would be part of what speakers would have to know in order to know that 10) is true. This is already quite implausible, since competent users of the term "slimy" don't have to have any particular convictions about the metaphysics of sliminess. But to make matters worse, even *if* the metaphysical theory were to enter into the semantics, the logical form--that is, the structure that feeds into compositional semantic interpretation-- of "Binky is slimy" would be:

GEN x,y[Standard Observer(x) \wedge Standard Condition(y)][ExperiencesAs_In_(x, Binky, slimy, y)]

The complexity of the clause makes it implausible that a child would learn how to interpret sentences structured this way before learning the much more simply structured clauses for quantifiers. More importantly, even if this were plausible, and there is a compelling argument for incorporating standard conditions and observers into the truth-conditions for 10), the formulation above delivers the wrong logical form for "Binky is slimy." The GEN operator is needed because the dispositional theory dictates that Binky is slimy iff Binky is experienced as slimy across a general *range* of conditions including more than one standard observer and more than one standard condition. But this is already problematic. "Binky is slimy" is a statement of a particular fact about a particular individual, and as such, is patently non-generic. Leslie thus concludes that a semantic

theory that takes us this far away from an empirically adequate account of the logical form and compositional structure of generics ought to be abandoned. The question I will pursue in the next few sections is whether this demarcation between the semantic theory for generics and the metaphysics of genericity holds up in the case of Type C generics.

3. Introducing Type C Generics

All Type C generics have two interesting properties. First, Type C generics predicate what I call *socially perspectival predicates* (SPPs) of social kinds. Second, by virtue of being generics, they represent social kinds as *non-accidentally* linked to the properties described by SPPs. We are here interested in a subset of Type C generics, namely those concerning racial kinds. Generics such as "Blacks are criminal/violent/thuggish" and "Muslims are terrorists" encourage *racialized* interpretations insofar as they lead interpreters to ascribe racial or ethnic characteristics to *criminality, violence, terrorism, etc.*

The rest of the chapter runs as follows. First, I introduce the central motivations of socially perspectival semantics and introduce the notion of a socially perspectival predicate (SPP). I then show that Leslie's weak semantics for generics and its attendant distinction between semantic and metaphysical truth can explain neither 1) how SPPs (and a fortiori Type C generics) acquire their truth conditional semantic features nor 2) how the underlying semantic structure of Type C generics make default racialized interpretations of these generics plausible. Finally, I sketch an account that explains both of these aspects.

3.1 Socially Perspectival Predicates

Throughout the dissertation, I propose that we need to understand many instances of racialized discourse as belonging to a much broader linguistic class which I will call *socially perspectival predicates*.

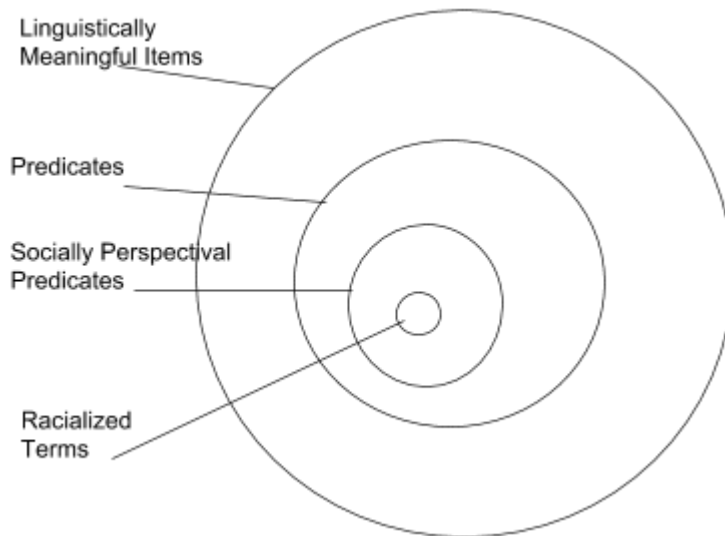


Figure 1

SPPs and Racialized Terms in the space of linguistic meaning

We will touch on the semantics of racialized terms at the end of this chapter, but much of what I want to say about the meaning of racialized terms is developed in Chapters 3 and 4. This chapter mainly focuses on getting clear on what “socially perspectival predicates” are, which is a lexical class which includes much more than just racialized terms. Understanding what is “social” and “perspectival” about these predicates will bring us closer to understanding how communication by means of racialized discourse works.

Let’s start by discussing the contrast class: non-perspectival predicates. According to received wisdom from a tradition going back to the very beginnings of Western philosophy, some of our concepts “carve nature at its joints.” Mountains, gold, water, and other mind-independent natural

kinds are the joints, and the terms or concepts for these joints are the knives. We may bring the *concepts* of “mountain” and “water” to the world, but those concepts are answerable to how mountains and water actually are. The world’s joints are metaphysically *perspective-independent* entities, and they are criterial for the correctness and applicability of terms/concepts which are *non-perspectival*.

Yet other concepts reflect the perspective of the categorizers, and indeed require such perspectives for their intelligibility. This is most obvious in the case of concepts for conventionally demarcated social kinds such as “foreigner.”¹⁵ Unlike the extension of a term like “mountain,” which always includes all and only those things in the mind-independent world which are mountains, the extension of “foreigner” changes depending which geopolitical perspective we honor. I can even experience the perspectival nature of “foreigner” by going to a foreign country and realizing that once there, all those foreigners think that *I* count as the foreigner!

"Foreigner" is an example of what I call a *socially perspectival predicate* (SPP.) A predicate P is an SPP iff it meets four interrelated conditions:

- 1) **Metaphysical condition:** the property or object X that P picks out is *perspective-dependent*, where perspective-dependence means a) that the *existence* of a perspective on X is necessary for the existence of X, and b) that the properties which make an object an X require the existence of a perspective.
- 2) **Pragmatic condition:** patterned socially sanctioned uses of P *institute* the property or object which P picks out, and facts about these properties constrain and direct these patterned uses.
- 3) **Epistemic condition:** there is no known social-practice-independent or non-conventional *epistemic* standard which can be used to determine whether a property or object X falls into the extension or anti-extension of P.
- 4) **Semantic condition:** P’s extension is partially fixed by a dominant or authoritative social perspective within the context of use.

¹⁵ I borrow the example of "foreigner" from Price 2005.

The metaphysical condition entails that the properties described by SPPs do not exist independently of the perspectives of the creatures who apply such predicates. Independently of labelling people as “thugs,” “criminals,” and “terrorists” there would be no thugs, criminals, or terrorists-- though there would be mountains, insects, quarks, and shapes, even if these objects were never labelled as such by a linguistic community.¹⁶ Moreover, the set of properties that are sufficient to render one a *criminal*, for instance, are similarly perspective-dependent-- assuming that breaking law L is sufficient to make one a criminal, for instance, one is a criminal from the perspective of some legislative institution which sanctions L-breakers. Of course, the same conduct would not be sufficient for being a criminal from the point of view of a legislative systems which does not sanction L in the first place.

The pragmatic condition suggests that dominant patterned uses of SPPs are capable of instituting or creating social kinds. There is a feedback loop between practices of labelling or describing bits of the world and facts about the world external to such practices. Under some conditions, labelling an object by use of P creates constraints on what kinds of things belong in P’s extension, and facts about these objects mutually constrain the proprieties of use for P.

The epistemic condition entails that the procedures for determining whether or not an object fits into the extension of an SPP are set by the judgments, conventions, and practices of a linguistic community, not by any convention-independent standard. In order to find out what sorts of people in a given society counted as *criminal*, *thuggish*, or *a terrorist*, we would have to make concrete reference to social conventions such as the criminal code of that society, norms of etiquette and proper conduct, and the overarching political context. Later, I will argue that pairing this

¹⁶ I take it that the existence of metaphysically perspective-dependent properties depend on the making of *actual* judgments about these properties, but I do not mean to foreclose the possibility that some such properties depend on there being *dispositions* to detect these properties, even if no perspective actually does detect them.

epistemic thesis with the metaphysical, semantic, and pragmatic conditions, yields the result that a community's *judgments* and *practices* are authoritative over what kinds of objects actually fit into the extension of a given SPP. As long as the community has fixed an interpretation of what it means to be a "criminal," it cannot be wrong about who counts as *criminal*.

These three conditions are compatible with the fact that the extensions of SPPs are frequently negotiated, re-legislated, and contested. These contestations arise when members of a collective disagree on certain aspects of the publicly shared criteria for categorizing an object or kind with the predicate. For instance, terms for human races such as "Black" and "White" are applied to different people according to a few loose criteria, including somatic characteristics (skin color, hair texture, etc.), prior ancestry, culturally-specific "Black" vs. "White" experiences, personal racial identification, etc. (Mills 1998) Very often, these criteria conflict when we try to categorize a given individual as White or Black (even assuming that we are only trying to categorize people of roughly European and/or African stock.) Borderline cases such as "mixed-race" individuals and individuals capable of "passing" as a member of another race show that these loose criteria deliver conflicting results depending on which aspects of the theory we regard as most relevant for any given case. In such cases, we might defer to a *meta-conventional* epistemic standard. For instance, one widely endorsed theory of racial membership--the "one drop of blood rule," according to which a certain percentage of "Black" ancestry is sufficient to make one "Black"--frequently settles debate about whether a person is White or Black in contexts in which that meta-convention is honored.¹⁷

¹⁷ One might think that the one-drop-of-blood rule represents a convention-independent epistemic standard for "Black" and "White," since what is being measured are non-perspective-dependent genetic facts of the matter. This is true, but it misses the point. The one-drop rule is a convention-independent epistemic standard which is invoked to justify a very much convention-dependent standard with a specific function: to subjugate those identified as "Blacks" and to privilege those identified as "Whites." It is simply not the case that the one-drop rule identifies two different racial groups independently of the activities of humans and communities which draw racial distinctions for specific purposes. For instance, a recent study (Bryc et al 2015) found that about 13% of Louisiana residents who identify as "White" have at

The basic ideas underpinning the metaphysical, pragmatics, and epistemic conditions are not new. As we saw in Chapter 1, the idea that humans can institute or construct a property, object, practice, or kind *F* just by regarding, treating, or describing things as *F-ish* is a familiar point from the philosophical literature on social construction. Yet this point is seldom seen as relevant to the *semantics* of the terms and concepts we use to identify *F*s. For instance, in a foundational text on generics, Krifka et al (1995) claim that any adequate semantics for generics must distinguish what generics say about “the world” and what they say about “cultural norms”:

For instance, suppose it is the norm in some culture to assume that snakes are slimy. Even in that culture, the sentence *snakes are slimy* is a false sentence-- although believed to be true by most members of the culture--since snakes, those real-world objects, are in fact not slimy. That is, generics are construed about making claims about the world, rather than what is considered a cultural norm. (Krifka et al 1995:49)

Krifka et al’s intuition pump is well-chosen: *of course* there is a mind-independent and language-independent fact of the matter about whether snakes are slimy, and a semantic theory of the truth of the sentence “snakes are slimy” need not concern itself with the idiosyncratic use of misinformed denizens of a particular culture. The more general assumption here is that a semantic theory of generics ought to show that generics describe the world as it is correctly, independently of our beliefs.

Yet this principle seems rather suspect when we consider generics that explicitly invoke social and normative concepts and categories, and especially those concepts or categories over which

least 1 "Black" ancestor over the last 11 generations (or 1% African ancestry.) The findings were hotly debated; did this make those white folks "Black?" Meanwhile, there was very little discussion about a related topic: the average African-American has 24% European ancestry, or at least one "White" ancestor over the last 3 or 4 generations. But no one seemed to wonder whether this made the average African-American "White." This would seem to confirm the existence of an implicitly agreed-upon conventional standard: one Black ancestor is sufficient to make you Black, but no amount of White ancestors is guaranteed to make you White.

the denizens of a culture can be *authoritative*. For instance, replace "snakes are slimy" in the above paragraph with "drugs are cool" and see how it stands-- "the sentence *drugs are cool* is a false sentence--although believed to be true by most members of the culture-- since drugs, those real-world objects, are in fact not cool." The fact that most members of a culture believe that snakes are slimy is neither a necessary nor sufficient condition for snakes being slimy, but the fact that most members of a culture believe that drugs are cool is necessary and possibly sufficient for it being the case that drugs are cool.¹⁸ One can easily think of other counterexamples to Krifka et al's requirement here, which is hampered all the more by the fact that it seems to rule out an analysis of virtually any generic concerning social or normative categories.

From this traditional point of view concerning the relationship between semantic theory and metaphysics, we can see that it is the *semantic* condition on SPPs that is most unorthodox. My claim is that despite appearances to the contrary, the world does and should intrude on our semantic theorizing. In particular, social perspectives which are dominant or authoritative affect the meanings of Type C generics. Yet showing this will require some work. First we will re-examine Leslie's weak semantics, and then see if we can motivate an alternative account of the semantics of Type C generics.

4. Leslie Reconsidered

If there are indeed predicates which are socially perspectival in the sense I've outlined here, Type C generics pose two problems for Leslie's account. First, they challenge her distinction

¹⁸ Another, perhaps more interesting example: "Cows are food," is true if "food" selects the dominant culinary-economic context in which cows are treated as things to be consumed, and the ability of "food" to select this context is entirely dependent on whether or not cows are treated as food by some set of dominant practices. At the same time, ethical vegans and vegetarians can, with some justification, argue that the sentence is false at other, less dominant contexts.

between semantic truth and metaphysical truth-makers by showing that in a wide range of cases, the things that must be the case in order for a generic to be true are sometimes the very conditions under which a predicate gets its semantic meaning. Those conditions frequently consist of *shared understandings* of what it is to *be* an X, and it is sometimes impossible to isolate these understandings from semantic questions about what "X" means. Second, the default racialized readings of Type C generics cannot be accounted for within Leslie's weak semantics.

4.1 Contra Leslie 1: Perspectives and Semantics

Leslie's reasons for keeping truth conditions separate from worldly truthmakers are good ones. Semantic theorizing about what the sentences of a natural language mean should be constrained by facts about how human beings manage to pull off the trick of learning that language in the first place. And if children are able to display adult-like competence with using and interpreting generics even before they show competence with quantifiers, that is good evidence for the compositional semantic profile of generics being very simple. However, what we have seen is that even if this profile is simple, the worldly conditions that must hold in order for a given generic to seem *true* to an interpreter are quite complex. In fact, they're too complex-- the clauses that have to be built into an analysis of worldly truth makers would take us far away from the original compositional structure of a sentence (recall our "Binky is slimy" example.)

However, this metaphysical/semantic distinction is not required for making sense of the semantic profile of generics. Here are two reasons why. First, Leslie's own theory is already committed to the idea that background knowledge substantially figures into an interpreter's approach to linguistic inputs. Leslie's truthmaker clauses specify not only what states of the world

would make a given generic intuitively true, but also the *epistemic* constraints under which an interpreter is laboring in order to properly interpret the sentence. Which generics an interpreter judges true will be a function of how that interpreter perceives worldly conditions to be. But more importantly, how that interpreter perceives those worldly conditions will depend on how or whether they grasp the *concepts* required for assigning a truth-value to the generic. For instance, Leslie's semantics predicts that a Type A generic such as "Police officers catch criminals" is true because criminal-catching is a salient "function or purpose" of police officers. Yet it *also* predicts that the generic will ring *false* to an interpreter who is ignorant of what it is police officers (should) do.

Second, the motivation for Leslie's semantic/metaphysical demarcation is undercut when we realize that we needn't choose between a compositional semantics for generics on the one hand and a theory which allows metaphysical, social, and epistemic constraints to inform semantics on the other. If generics indeed are a source of evidence for how our primitive generalization mechanisms work, then it is plausible to think that interpreters privilege different properties of kinds in assessing certain generics. For instance, the following generics are true, and the type of facts that seem to make them true (to an interpreter) differ in each case:

- 11) Turtles are long-lived. (*biological endowment*)
- 12) Dobermans have pointy ears. (*artifactual*)
- 13) Dobermans have floppy ears. (*natural/ non-artifactual*)
- 14) Scotsmen wear kilts. (*social/ cultural tradition*)
- 15) Prime numbers are divisible by one and themselves. (*axiomatic/ definitional*)

Children and adults alike treat some types of worldly properties rather than others as more important to the evaluation of a given generic, depending on what specific kinds and ascribed

properties are involved. And if these metaphysical properties have such a strong effect on whether a generic is true or false, then that seems to be sufficient reason to incorporate *metaphysical* theories of what makes a generic true into the semantics for the generic.¹⁹

To this extent, Leslie's worry about compositionality loses some of its bite. Even if it is implausible that children are tiny metaphysicians, fully equipped with dispositional theories of sliminess, that does not show that children are completely ignorant of what kinds of worldly features make it the case that a generic is true. All we need in order to treat the metaphysics as reasonably continuous with the semantics is the premise that children rely on their knowledge of the world in interpreting and assessing generic claims,²⁰ and that they frequently defer to other members of their linguistic community in assigning meanings to terms.

The question, then, is what kinds of properties are relevant to the interpretation of the Type C generics we are tracking? A great deal will hinge on what properties are relevant to the interpretation of various SPPs. In the remainder of the paper, I motivate the claim that in order to make sense of the interpretation of SPPs like "terrorist," "criminal," and "thug," we need to allow the semantic dimensions of SPPs to interact with facts about their function within particular sets of social practices.

¹⁹ For instance, Gelman (2003) shows that children are far more likely to produce generics about "natural" kinds than "artifactual" kinds. She concludes that children rely on the deliverances of an implicit folk metaphysics in evaluating and producing generic claims. Prasada and Dillingham (2006) show that adults access different explanations of why a generic is true depending on their folk conception of what ontological class objects and kinds belong to. While this does not conclusively demonstrate that such "folk theories" have anything to do with the semantics of the terms implicated within them, it does suggest that our competence with semantic values reflects a messy acquisitional process in which conceptual boundaries are perhaps not as distinct as mainstream semantics seems to suppose.

²⁰ See Prinz (2002) for a thorough defense of the claim that background knowledge can figure into semantic interpretation in a way which still obeys the compositionality constraint. I must remain agnostic on what counts as metaphysics being "reasonably continuous" with semantics for the moment, but if Gelman's and Prinz's acquisition stories are workable in the case of social kinds, we have good grounds to challenge Leslie's dichotomy.

4.2 *Contra Leslie 2: Racialized Interpretations and Weak Semantics*

Leslie's semantic clauses are also supposed to explain why it is children *and* adults have such strong semantic intuitions about the truth of generics like "Turtles live long lives" and "Mosquitoes carry West Nile" while having equally strong intuitions about the falsity of generics like "Ducks are female" or "Bees are sterile." Consequently, the notion of racialized interpretation gives Leslie's weak semantics a test: if her account can successfully explain why certain default interpretations are characteristic of conventionally racialized SPPs, then it becomes less clear that we ought to explain facts about how generics are interpreted in terms of social perspectives. However, I contend that her semantics fails this test for the same reasons as before: in treating semantic truth and metaphysical truth-makers as isolated from one another, Leslie's semantics cannot give an adequate account of either the meaning or metaphysical truth makers of Type C generics with racialized SPPs.

To see this, let's revisit Leslie's semantics for Type B generics, which is organized around the following "striking property clause."²¹

Striking Property Clause: If F is a striking property, then some K's are F, and all other K's are disposed to be F. If F is not a striking property, [and if "K's are F" is not a minority Type A generic,] then almost all K's are F.

Type B generics thus admit of the following paraphrase:

16) Mosquitoes carry West Nile.

16a.) Some mosquitoes carry West Nile, and all are disposed to carry West Nile.

²¹ One might think that Leslie accounts for these readings via her "social kind" clause: "if K is a social kind, then F is the *function or purpose* of the kind K." However, I think it implausible, to say the least, that *criminality* or *terrorism* are primarily conceived as (much less actually are) the function or purpose of blacks and Muslims. If such things were thought to be functions or purposes of these populations, for instance, it would be difficult to *fault* such kinds *for* performing those functions (although one would still be able to argue that those functions are bad.) In any case, there is evidence (Prasada and Dillingham 2006) that social kind generics are not normally given this teleological interpretation.

Yet such paraphrases do not adequately capture the nature of the relationship that obtains between kinds and predicates in some Type C generics. Consider:

17) Blacks are criminal.

17a.) Some Blacks are criminal, and all Blacks are disposed to be criminal.

18) Muslims are terrorists.

18a.) Some Muslims are terrorists, and all Muslims are disposed to be terrorists.

If racialized interpretations are as accessible as I have suggested, these paraphrases are far too weak, regardless of whether they are supposed to be "semantic" or "metaphysical" clauses. Assume that they are metaphysical truth-makers. Thus the "some" paraphrase is supposed to specify what the world must be like if an interpreter is to judge the generic true. Yet this would not explain why 17) and 18) seem to exert some inductive or cognitive "pull" on many interpreters while 19) and 20) do not seem to exert such a pull.²²

19) Some lawyers are criminal, and all lawyers are disposed to be criminal.

20) Some Swiss people are terrorists, and all Swiss people are disposed to be terrorists.

²²Leslie (2017) has argued convincingly that generics like "Blacks are criminal" and "Muslims are terrorists" are structurally similar to Type B generics, at least in their psychological function (that is, these generics are closely linked to essentializing, stereotyping, and prejudicial attitudes.) This chapter argues that Leslie's semantic glosses for many of the generics she countenances as Type B generics are insufficiently rich to account for the target perspectival vocabulary on display in the case of Type C generics (hence the need for a distinction between Type B and Type C generics.) However, as will become apparent in Chapter 6, I am happy to treat Type C and Type B generics as very closely related in their respective *psychological* functions. As we will see, both sorts of generics are both reflections of and vehicles for essentializing attitudes, and may have similar effects on the flow of information among interlocutors.

The cognitive "pull" in question is that interpreters are likely to rely on representations of Blacks as criminal and Muslims as terrorists even despite few to no negative experiences with blacks or Muslims. The majority of Americans grossly overestimate the objective share of crimes committed by Blacks (Blow 2014), as well as the share of terrorist activity committed by Muslims (Ahmed 2015), and there is evidence that interpreters only need very scant evidence for Black criminality or Muslim terrorism in order to tacitly endorse a generalization to the effect that "Blacks are criminal" or "Muslims are terrorists" (Leslie 2017). That is, Blacks and Muslims are perceived to be *more disposed* toward criminality and terrorism than, say, lawyers and Swiss people.

However, these paraphrases do not explain why it seems that the evidential bar for accepting "Blacks are criminal" is so low, and why the bar for accepting "lawyers are criminal" is so high. After all, the objective prevalence of crime among lawyers is comparable (and perhaps greater) than the prevalence of crime among Blacks, but presumably crime among Blacks is given the greater weight. My claim is that we can explain this disparity by hypothesizing that the term "criminal" takes different semantic values in each sentence-- to an interpreter, the semantic threshold required for *Blacks* to belong in the extension of "criminal" is lower than the threshold required for *lawyers* to belong in its extension.

Of course, Leslie might argue that we are not forced to a *semantic* explanation of this disparity at all. The reason that "Blacks are criminal" is more readily accepted than "Lawyers are criminal" is simply that the stereotype linking Blackness and criminality is inductively more potent, or more widely endorsed, than the stereotype linking lawyers and criminality. In other words, the job of her metaphysical truthmakers is simply to specify minimal necessary conditions which must be met in order for the generic to be judged true, and specifying any interpreter biases are not the job of her

conditions (though it might be the job of a psychological theory of the mechanisms underlying selective stereotyping, bias, and prejudice.) In fact, all we need to explain why Type C generics are accepted so readily is that interpreters simply tend to overestimate the prevalence of criminality among Blacks, in the same way that they overestimate how many sharks attack swimmers. Why is this not explanation enough?

In response, recall that Leslie's "wordly truth-makers" are supposed to provide a partial explanation of why interpreters accept the generics they do. In the case of Type A generics, the fact that there are actually regimented similarities among bird-reproduction explains why "Birds lay eggs" is generally accepted. In the case of Type B generics, the fact that "attacking swimmers" or "carrying Lyme disease" generally strikes humans as an alarming or dangerous property explains why they are prone to overestimate the prevalence of swimmer-attacking and Lyme-disease-carrying among sharks and ticks, respectively. But again, *egg-laying*, *swimmer-attacking*, and *Lyme-disease-carrying* are not in general perspective-dependent properties, in my sense. Interpreting these sentences does not require mastery or familiarity with a social perspective which is authoritative over the extension of these predicates (although it does of course require significant world-knowledge.) On the other hand, as I will argue in the next section, the perspective-dependent facts about who counts as a *terrorist* in a social structure, and the procedures for identifying terrorists, *are* features of the world which are relevant to determining the conditions under which interpreters assign semantic interpretations to Type C generics. That is, they are actual or perceived regularities which should play a role in explaining why interpreters assign the interpretations to sentences that they in fact assign. Consequently these are precisely the sorts of things that should be included in Leslie's clauses, by her own lights. "Muslims are terrorists" isn't judged true only because "some Muslims" are terrorists;

rather, it is judged true because its semantic interpretation (and perceived truth-conditions) is sensitive to a race-specific dominant social interpretation of what it is to *be* a terrorist given a set of ostensibly normal social conditions.

Finally, note that treating these paraphrases as giving the *truth conditions* of Type C generics rather than their truth-makers (contrary to Leslie's intentions) does not get around this problem. The trouble is that the "some" paraphrase remains true under virtually any substitution of a social kind; as long as one member of the kind has the property ascribed, the sentence is true. The paraphrases would thus obscure the sense in which being criminal or being a terrorist is perceived to be *proper* to certain kinds rather than others.

4.3 Backing things up

Of course, these objections are in some sense mere promissory notes. After all, I haven't yet shown that socially perspectival predicates are a real lexical type, nor that this framework is preferable to Leslie's in making sense of generic meaning. So at this point, the problem for Leslie is provisional-- *if* we can show that a perspectival semantics is viable for Type C generics, we have stronger evidence that the two problems I have posed for Leslie stick.

5. Socially Perspectival Semantics

This rest of the chapter has two goals. The first goal is to sketch an alternative account of linguistic meaning which gives pride of place to the role of social perspectives. The second is to use this framework to sketch an alternative account of Type C generics. Meeting the second goal allows

us to i) adequately model the lexical context-sensitivity of Type C generics, and ii) to explain why racialized interpretations are so accessible for Type C generics.

We will begin with an argument for thinking that the sorts of lexical items I have called socially perspectival predicates constitute an interesting linguistic type. Then we will move to a defense of the semantic condition on socially perspectival predicates:

Semantic condition: P's extension is partially fixed by dominant or authoritative social perspectives within the context of use.

Making sense of the semantic condition requires us to answer three questions:

- How do social perspectives become dominant or authoritative?
- What role do social perspectives play in online interpretation?
- What is a social perspective?

Finally, we will be in a position to give an alternate account of the semantics of Type C generics.

5.1 Proof of Concept: Baseball

In order to show that SPPs have the type of lexical complexity I am supposing they have, we need to show two things. First: that there are predicates whose extensions are determined by social, conventional, practice-dependent standards, and that those standards yield *authoritative* interpretations of the extension of those predicates. Second: this process of extension-fixing has “metaphysical” effects. Namely, in setting the extension of the predicate, a *kind of object* is instituted.

Let's start simple. Imagine a game of baseball. The game is governed by two kinds of rules, *constitutive* and *regulative*.²³ Constitutive rules for a game or practice are rules which somehow

²³ The basic contrast between these types of rule has been unpacked in different ways by Rawls (1955), Searle (1995), and many others.

"constitute" a given game or practice. The constitutive rules of baseball make it the case that baseball games can exist: these rules specify what counts as a ball or strike, under what conditions a team earns a run, and that if a batted ball is caught, then the batter is out, etc. Regulative rules, on the other hand, "regulate" an already existing practice: they specify appropriate rules of conduct for the players. Thus a batter who strikes out is obliged to leave the field and sit in the dugout, and a runner is obliged to run the bases if he is to score a run for his team. Obviously, constitutive and regulative rules interact. A player who conducts herself appropriately also manages to follow the constitutive rules of the game.

The rules of baseball manage to confer semantic contents on linguistic performances within the game and *also* manage to constitute baseball objects. Deciding just *which* baseball objects they are is largely a matter of what perspectives we are prepared to treat as authoritative. Let's take the example of strikes. When is a pitch a strike? There are at least three natural interpretations, each of which has substantial effects on how the game is played.

Let X be a pitch from pitcher to catcher.

Let Condition S be met just in case X passes through the *strike zone* (**Figure 2**).

Let U be the home-plate umpire. U's job is to respond to any X with one of two calls: "ball!" or "strike!" Any X that is not a ball is a strike, and vice versa.

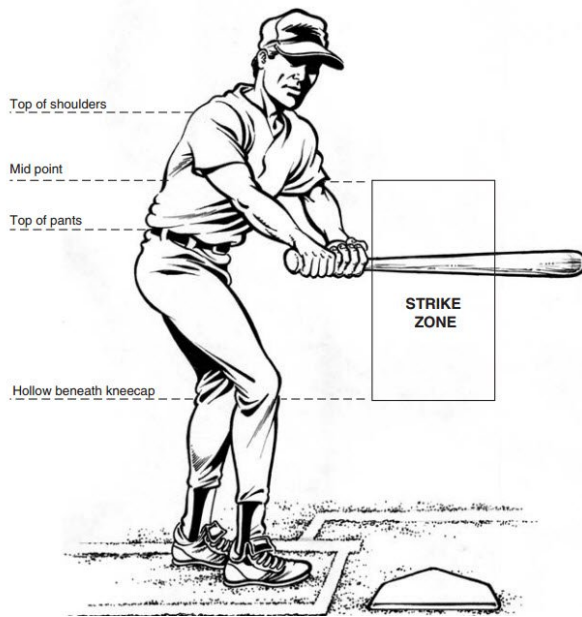


Figure 2

Source: Official Major League Baseball Rules, 2018

STRIKE₁: X counts as a STRIKE₁ just in case X meets condition S, regardless of whether or not U calls "strike!" U's call is in a sense entirely superfluous.

STRIKE₂: X counts as a STRIKE₂ just in case condition S is satisfied *or* U calls "strike!" U's calling "strike!" is sufficient for X to count as a STRIKE₂, but it is not necessary. That is, it is possible for an X which satisfies condition S to count as a STRIKE₂, even if U does not call "strike!"

STRIKE₃: X counts as a STRIKE₃ just in case U calls "strike!," regardless of whether condition S is met. U's calling "strike!" is necessary and sufficient for X to be a STRIKE₃.

Each idealized interpretation has different implications for the epistemic authority that U's call bears.

For instance, if what it is to be a *strike* is to be an X of type STRIKE₁, U can be wrong, even systematically wrong, about whether any given X "really" is a strike. In fact, U's call is entirely superfluous. Strictly speaking, U doesn't play any special role in conducting the game. The question of whether X is a strike could be resolved by appeal to a replay device that measured precisely whether or not X satisfied condition S. At the other extreme, if what it is to be a *strike* is to be an X

of type STRIKE₃, then it is *impossible* for an umpire to be wrong about whether a given X is a strike. His utterance, "strike!" functions as a performative which *makes* it the case that X is a strike, much in the same way that saying "I now pronounce you married" in the right context makes it the case that two people are now married. At this extreme, it does not even seem as if U is still bound by the constitutive rules: rather, he legislates them as he goes along.

We need an intermediate position which does justice to the notion that the umpire's call is answerable to how things stand in the world (that is, with the facts about X's) and to the notion that the umpire's call is authoritative over what those X's actually are (i.e. balls or strikes.) STRIKE₂ is just such a position. According to STRIKE₂, an umpire can be wrong *or* right about whether any given X is a strike, depending on which perspective we treat as authoritative. Say that an umpire calls an X that does not satisfy condition S. From the point of view of the constitutive rules laid out by the official rulebook, U is clearly incorrect about whether X was *really* a strike. However, from the point of view of the regulative rules, which specify how players and umpires are supposed to conduct themselves during a baseball game, the umpire's call is authoritative. As it turns out, the regulative rules stipulate that once an umpire makes a strike call, that judgment can be neither challenged nor overturned. The practice of strike-calling is thus functionally *oligarchical* -- every player on the field and every spectator could deny that the umpire's call was correct, but that wouldn't make the sentence "that was a strike!" any less true.²⁴ In short, in baseball, the umpire's making a *mistake* in calling X a strike (according to the rulebook perspective) is completely compatible with X's *being* a strike (according to the perspective honored in practice.) This, I submit, is good evidence that "strike" is a perspectival predicate.

²⁴ Of course, the perspectives of authoritative bodies that outrank the individual umpire-- the Major League Umpire's Union, say, or Major League Baseball-- might reverse this valence.

“But wait,” you might say, “why assume that the *truth value* of ‘strike!’ can vary with which perspective is treated as relevant? I can see at least two ways in which this delivers the wrong result. First, you haven’t given me an independent reason to treat the umpire’s perspective as *semantically* relevant. Sure, it’s relevant to the kinematics of the game--we can’t proceed unless the umpire makes *some* call-- but that’s not sufficient for showing that the umpire’s call is relevant to the *semantics* of “strike.”

“Second, you are assuming that authority over how the game proceeds gives the umpire a form of authority that is relevant to the semantics of ‘strike,’ but at best, umpires have an *epistemically* privileged position when it comes to calling strikes. Why not assume that the umpire is simply getting things wrong when he calls strikes which do not satisfy the rulebook notion of strike, full stop? That is, the umpire knows what ‘strike’ means-- it just means what the constitutive rules say-- but he is frequently mistaken about what kinds of things actually are strikes, much in the same way that I can misperceive a cow *as* a horse, even if I’m an expert on cow-identification. In fact, ‘strike’ calls *must* have merely epistemic (and not semantic) significance. After all, umpires can only make so many mistakes until they are sanctioned by Major League Baseball. It’s hard to see why such sanctions would be appropriate if the umpire speaks *truly* every time he calls a strike.”

I believe that these objections ultimately arise from an insufficient appreciation for how social practices among competent human language users can confer *functional* semantic properties on the terms of a language. Sally Haslanger’s (1995, 2007) illuminating vocabulary of "manifest" and "operative" concepts is helpful in understanding the basic idea. For Haslanger, a "manifest" concept is explicit, public, and officially recognized. Applying manifest concepts is ultimately a casuistical problem. In applying manifest concepts, we need to appeal to a more or less explicit codification in

order to subsume a given part of the world under a concept. In the context of baseball, then, the manifest concept of a "strike" is just what the official rule book says that it is. On the other hand, an "operative" concept is more implicit; it is something of an "unwritten norm" which is practiced despite its lack of official institutionalization. STRIKE₂ best captures the operative notion of a strike: while the umpire can be wrong about what a strike is from the perspective of the manifest interpretation, he cannot be wrong according to the operative interpretation--he makes the call, and the game goes on. In this case, it is the functional, operative notion that is privileged within the practice of baseball, even though it is clear what it is for any X to be a strike from the point of view of the manifest concept.²⁵

Second, the objection above does not so much argue for as *presuppose* that strikes are best thought of as perspective-independent entities--that is, that there is a perspective-independent *fact* of the matter about whether or not the ball crossed some part of the plate, and thus we should draw a distinction between strike-seemings and actual strikes. However, this misses the point of the example: it's not that there are two notions of "strike" and that one is a perspectival notion, it's that there are (at least) two notions of "strike" and that *both* are perspectival. The argument is over *which* of these perspectives we should treat as relevant-- the rulebook perspective or the umpire perspective? I have argued that the umpire perspective is the one which is semantically and epistemically relevant because of features of how baseball actually works in practice.

²⁵ One might well wonder why the operative interpretation of strikehood is not STRIKE₃, according to which the "strike" call functions as a performative which makes it the case that a given X counts as a strike, just as "you are now married," expressed in the right kind of circumstance, makes it the case that a given X and Y are married. The reason that this seems wrong is that this would entail that the constitutive rules play *no* role in constraining what kinds of X's can be strikes: an umpire can be systematically and utterly wrong according to the constitutive rules and still not be able to be sanctioned. However, this seems to represent a breakdown in baseball normal conditions. Major League Baseball seems to agree: they sanction their umpires if they are *systematically* out of step with what the rulebook specifies.

Finally, the fact that a perspective is "in play" does not entail that everyone (or indeed anyone!) participating in the baseball game *consciously adopts* this perspective themselves, or knows that this is the perspective that is governing some aspect of game. My account makes room for the possibility that even the umpire *himself* is not authoritative over which perspective is in play. In fact, the umpire, no doubt, thinks of himself as trafficking in the manifest concept-- he takes himself to speak from the rulebook perspective. An umpire might even admit that he made a mistake after the game ends, suggesting that the umpire takes himself to be honoring the rulebook perspective, and expresses regret at failing to live up to it. However, on the externalist-friendly picture we've set up, features of how baseball is played in non-ideal circumstances may be different from what baseball practitioners *take* to be the authoritative perspective. In this sense, the umpire perspective is authoritative *even when the umpire does not recognize it as his own perspective*. This possible mismatch between *acknowledged* speaker commitments on the one hand and *actual* speaker commitments on the other is especially important when we turn to an investigation of how racialized terms function within discursive practices-- despite one's intentions to the contrary, it is possible to signal allegiance to a dominant perspective simply through the use of a word.

5.2 Racial Perspectives

What the earlier objection gets right, I think, is that within a highly structured social practice like baseball, the choice of the perspective that is relevant to determining the semantic and metaphysics of "strike"/*strike* is more or less arbitrary. My diagnosis is compatible with an alternative account that treats rulebook perspectives as authoritative. Arguably, that's because in baseball, *nothing*

much hinges on what perspective we treat as relevant.²⁶ However, nailing down the perspective which determines the functional meanings of our racialized terms matters a great deal for racialized ontology. Why? Because the functional properties of racialized talk very well *can* play a global role in organizing racial populations in a way that is highly consequential, and perhaps even dominating and unjust. Here the stakes are much higher, and the question of which perspective we treat as semantically and ontologically relevant is more important.

While racism is surely more complex than baseball, both are highly norm-governed systems, and both impose interpretations and meanings on the objects within their domain.²⁷ For instance, both the theorists and folk of oppressed groups have long known that dominant epistemic perspectives can dictate not only what sorts of entities count as *persons*, but also what different classes of *people are* in a functional sense. Consider even non-explicitly racialized terms like "personhood," "citizenship," and "equality." The famous opening words of the American Declaration of Independence, arguably the cornerstone of much of our manifest, official, publicly avowed thinking about human dignity, frame these categories as follows:

We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty and the pursuit of Happiness. (Jefferson/Franklin 1776)

On the other hand, Blacks and other groups have long held privileged perspectives on what concepts such as *personhood*, *justice*, and *dignity* mean in their practiced, operative sense:

When White people say "justice," they mean "just us." (Black American folk saying, quoted in Mills 1998)

²⁶ Or at least it doesn't matter in the sense of being highly relevant to ethical and social concerns *outside* the practice of baseball. The choice of baseball perspective of course does matter a great deal in the local world of high-stakes baseball contexts (no-hitters, perfect games, the World Series, etc.)

²⁷ McGowan (2012) especially draws attention to the notion that racism is normatively structured in her discussion of hate speech, and Shelby (2003) argues that racist ideologies have a cohesive normative structure.

Race was far from irrelevant to [the Enlightenment conception of] personhood; skin color, hair, and facial features were used to categorize people and determine their moral standing... that is the unacknowledged dark side of the Enlightenment ideal...All persons are equal, but only White men are persons. (Mills 1998)

The lesson generalizes. When we reflect on the manifest uses of our concepts, we find that there are a lot of things we know about the social world, and a lot of things we agree upon. This comprises a rich shared stock of social knowledge. We know that anyone can be a *criminal* or *terrorist*, given the right circumstances. We know that every person in our society is entitled to life, liberty, and the pursuit of happiness. However, reflection on the operative dimensions of SPPs like "criminal," "terrorist," and "person" tells another story. While anyone *can* be a criminal, Blacks are more criminal than Whites. While all persons are equal, Blacks are less equal than Whites.²⁸ While anyone can be a terrorist, the real terrorists are motivated by a fundamentalist interpretation of Islam.

Our stock of shared, explicit, "manifest" knowledge has a murky underbelly. We implicitly rely on culturally distributed representations, beliefs, and assumptions that structure our interpretation of the social world and shape our social practices, sometimes resulting in profoundly unjust social arrangements. Many of these representations are racialized, and seep into terms in our language, thereby becoming associated with the conventional meanings of terms like "thug," "criminal," and "terrorist."

How do racialized terms come to mean what they do, and how do we manage to have this implicit knowledge about who and what fits into the extensions of racialized terms? Over the next

²⁸ In fact, the operative and manifest notions of personhood came into sharp concordance during the "Three Fifths Compromise" of 1787, which treated each slave as 3/5ths of a person in tallying the total population of a state in order to assign the appropriate number of representatives. Southern states thus received more representation than they would have had had slaves not been counted. The result was an overrepresentation of slaveholder interests in Congress until 1865.

few chapters, we develop at least the beginnings of an answer to this question.

Our baseball example suggests that we ought to be attuned to the ways in which *power* and *authority* function within discursive practices. In practices like baseball, which perspectives are relevant and authoritative is determined partially by the rules of the game (e.g. the rules stipulate that the umpire is the authority) and partially by the way that the practice plays out in non-ideal circumstances (e.g. the relevant, authoritative perspective is determined by a combination of the umpire's perspective and the "operative" perspective that is observed in practice by the game's participants.)

Things are not so clear in more complicated discursive contexts, however. Partially this is because there are seldom explicit authority-conferring rules which make clear what perspective is in play. But a second aspect of the problem is that even when it is clear that a dominant social perspective has taken hold of a region of linguistic practice, it seems to mask the mechanisms that are responsible for creating and reproducing it.²⁹ The processes that reproduce dominant perspectives are often "loopy" in the sense that top-down authoritative mechanisms specify and reinforce the relevant perspective, and bottom-up deferential patterns become reflected in large scale institutional structures. For instance, the relevant perspective that determined membership within the political category of *personhood* is not solely the result of the "top-down" legislation that explicitly denied political and moral standing to non-Whites and non-males, nor is it solely the result of the "bottom-up" individual prejudices and material interests of powerful, landowning Whites. Rather,

²⁹ As many readers will note, this is also a feature of ideologies more generally. To be clear, I have very little to say about ideology in this dissertation, for two reasons: 1) I do not want to get bogged down in terminological or substantive issues on the points of similarities and difference between "discourse" and "ideology" (though for an illuminating comparison, see Purvis and Hunt 1993). 2) I do not want to wade into internecine debates within the theory of ideology, such as the debate about whether the theory of ideology is dedicated to a "descriptive" or "normative" project, or whether ideologies are *fundamentally* representational belief systems, social practices, or something else.

each of these forces (and others) played a role in the *constitution* of a perspective from which the extension of "personhood" could be determined. Moreover, the perspective that was so constituted was highly cross-contextual: Blacks were not only non-persons in most legal contexts, but also in most economic, social, and political contexts. We now turn to the question of what social perspectives are, and how they can be accessed in online interpretation.

5.3 *A Characterization of Perspectives*

Saying that the semantics of baseball vocabulary and racialized vocabulary is sensitive to the deliverances of dominant social perspectives is one thing. Yet we also might want to know what these sorts of perspectives fundamentally *are*, and how they manage to determine the content of perspectival terms. We'll take each of these in turn.

What is a perspective, social or otherwise? My conception of a perspective largely follows Elisabeth Camp's (2013) illuminating characterization:

Above all, perspectives are ongoing dispositions to structure one's thoughts, along at least two dimensions. First, a perspective involves dispositions to notice and remember certain types of features rather than others, so that those features are more *prominent* or salient in one's intuitive thinking, and have more influence in determining one's classifications. Second, a perspective involves dispositions to treat some classes of features as more *central* than others, in the sense of taking those features to cause, motivate, or otherwise explain many others. (Camp 2013: 335-336)

Perspectives thus have descriptive, explanatory, and evaluative dimensions. On the descriptive level, perspectives have a "foreground-background" structure. If I want you to see the way my living room looks from a certain perspective, I can position you in a certain sightline, tell you to attend to certain things, and in general influence your location and attention in ways that make "foreground" or

“central” objects seem highly relevant (the light hitting the coffee table, the books on the shelf) and “background” or “peripheral” objects seem less relevant (the cobwebs in the corner.) Yet beyond this, perspectives also contain information about what sorts of things *explain* the fact that things “look” a certain way (e.g. *why* is the coffee table part of the foreground?) Finally, perspectives involve evaluation insofar as a given perspective represents some features as more important to the identity of the things they represent, and others as less important. These three dimensions also characterize linguistically relevant perspectives.

5.4 Perspectives and Linguistic Meaning

Now, what role might perspectives play in the determination of linguistic meaning? Again, we’ll start with a toy example. Consider a sentence involving the preposition “to the left of”:

21) The man is to the left of the cat.

What proposition does this sentence express? Perhaps the most natural truth-conditions can be represented as follows:

22) 😊 🐱

Yet it is also possible that the sentence expresses:

23) 🐱 😊

Assuming that none of the terms in the sentence are ambiguous, how is it possible for the sentence to express two propositions which seem to contradict one another? The answer, of course, is that the sentence expresses any of a range of propositions depending on the perspective from which the

spatial relation is being reported. 22) is true according to a perspective which faces the man and the cat, and 23) is true according to a perspective located “behind” the man and the cat. (Interestingly, 23), but not 22), also gives the correct truth-conditions of the sentence from the *cat’s* spatial perspective.) The example suggests that spatial phrases like “to the left of” are perspectival. Their contribution to truth-conditional content can systematically shift depending on which perspective is being considered.

Of course, this example still leaves it open just *how* “to the left of” is perspectival. In general, perhaps the most direct way to accommodate perspectives within lexical semantics would be to argue that many lexical items, and all SPPs, contain an implicit argument for a social perspective. In fact, I have endorsed this view in the past (O’Donnell 2017). However, it seems to me now that this approach is confronted by insurmountable difficulties.³⁰ Consequently, I suggest that social perspectives play a *metasemantic* role in determining the meaning of racialized terms.

It’s worth dwelling on the distinction between semantics and metasemantics for a moment. It is common practice in philosophy of language to distinguish questions about *what expressions mean* from questions about *what makes expressions mean what they do* (Dummett 1974, Kaplan 1989, Stalnaker 1997.) The former sort of questions are “semantic” questions insofar as they are questions about which semantic values (meanings) are paired with which expressions. The latter sort of questions, however, are “metasemantic” questions, insofar as they ask for an explanation for *why or how* a given expression is paired with a particular semantic value.

³⁰ In short, the core problems with this “semantic-level approach” are that it i) lacks evidential support from syntax and ii) that it likely trivializes the core thesis of semantic perspectivalism. However, the semantic-level approach works quite nicely for perspectival vocabulary which is not necessarily *socially* perspectival in my sense. In the interest of minimizing distractions, my reasons for attraction to and subsequent disillusionment with this approach are disclosed in Appendix 1.

More precisely, while semantics strives to give an account of what the meaning of term T and what its contribution to truth-conditional content might be, metasemantic theories investigate the causal, historical, psychological, logical, and other mechanisms which explain *how or why* it is that T has the semantic value that it does.³¹ So, to take an example, imagine that you think that the meaning of a proper name like “Madagascar” is identical to the term’s referent. That’s a (Millian) semantic theory of proper names, and one that would entail that the meaning of “Madagascar” just *is* a unique individual, namely the island nation off the coast of Southeast Africa. However, you might also be interested in *how* “Madagascar” came to refer to that unique individual island. Presumably, this would require us to do some historical research into how Madagascar acquired its name (apparently through a mistransliteration of “Mogadishu” by famed but confused Italian explorer Marco Polo in the 13th-century), and how that name became the standard way to refer to the island (presumably through semantic deference to Marco Polo, the various ways in which [mis]information can be spread, etc.) All this belongs to a metasemantic theory of proper names in general, and of “Madagascar” in particular.

So, to say that social perspectives play a metasemantic role is to say that ground explanations of why perspectival terms have the semantic values that they have. On this view, social perspectives *indirectly* fix the semantic values of racialized terms by fixing the semantic values of whatever context-sensitive elements can be found in the semantic profile of those terms.³² Social perspectives

³¹ This characterization clashes a bit with Burgess and Sherman’s (2014) decision to “take the central modes of explanation in metasemantics to be *metaphysical* (even where its explanantia or explananda turn out to be epistemic or psychological.)” I think it’s plausible to think that there are properly metasemantic projects which do not reduce to a discussion couched in the philosopher’s jargon of “grounding,” “supervenience,” etc., but this issue won’t occupy my attention here.

³² As mentioned at the beginning of the chapter, I actually think that metasemantic perspectivalism is general enough to account for many other regions of discourse. In other work, I argue that the framework can be adapted to model the semantics of terms for character traits such as “generous,” “cruel,” “friendly,” “cowardly,” etc.

need not be linguistically encoded in terms as contextual parameters. Rather, perspectives indirectly determine the semantic values of SPPs by fixing the values of whatever contextual elements are included *within* those SPPs.

6. The Semantics of (Some) Type C Generics

Now we are in a position to give an account of the semantics of at least some Type C generics. Let's begin by considering Tommie Shelby's characterization of anti-Black racist ideology:

Thus the view that blacks are inherently of low intelligence is part of a wide-ranging and interconnected set of beliefs that includes, on the one hand, beliefs about the laziness, aggressiveness, and unreliability of black people, and on the other, beliefs about their natural musicality, athletic talent, and sexual prowess. (Shelby 2003: 159)

Presumably, interpreters who are in the grip of anti-Black racist ideology explicitly or implicitly accept generics such as “Blacks are lazy,” “Blacks are aggressive,” “Blacks are unreliable,” “Blacks are unintelligent,” as well as more positively valenced generics such as “Blacks are musical,” “Blacks are athletic,” and “Blacks are sexually competent.”³³

The reasons for focusing on these sorts of generics are twofold. First, as Shelby claims, they are cornerstone tenets of anti-Black racist ideology. Second, all of these generics predicate *gradable adjectives* of Blacks. We focus on them because they give us a straightforward way of modelling the metasemantic contribution of social perspectives.

Gradable adjectives can be paired with degree modifiers like *more*, *very*, *somewhat*, *kind of*, etc.:

24) Tim is very tall.

25) Alicia is kind of smart.

³³ Nickel (ms) explicitly connects racist ideology endorsement to the acceptance of certain generics.

26) Murdering someone is more criminal than stealing.

According to one orthodox analysis (Kennedy 2007, Glanzberg 2007), gradable adjectives map the objects that they apply to onto *scales* relevant to the given gradable. So, for instance, “tall” in 24), when modified by “very,” maps Tim onto the “upper end” of the scale for height, whereas “smart” in 25), when modified by “kind of” maps Alicia onto the more-or-less “middle range” for intelligence.

The comparative form of gradable adjectives (e.g. “lazier” ‘smarter,’ etc.) can thus be represented as a function which maps an individual object or kind onto some degree of tallness, criminality, or intelligence that is greater than the degree of tallness, criminality, or intelligence possessed by some other individual or kind:

27) a. Tim is taller than Alicia.

b. The degree of height of Tim is tall is greater than the degree of height of Alicia.

c. $\text{height}(\text{Tim}) > \text{height}(\text{Alicia})$

In the absence of an explicit comparison class (i.e. in a sentence like “Tim is tall”), the meaning of the positive form of a gradable adjective is a function from individuals to some degree that *exceeds some contextually determined standard for height*, $s(\text{height})$:

28) a. Tim is tall.

b. $\text{height}(\text{Tim}) > s(\text{height})$

Of course, since the semantic value that s takes is determined by the context, 28)a. might be true when Tim is in a room full of kindergarteners and false when Tim is at a Richard Kiel lookalike competition. There are other mechanisms by which these standards shift as well. For instance, a

general feature of gradable adjectives is that what standard seems (to an interpreter) to be most contextually appropriate can vary depending on what sorts of objects G is describing. The standard an object must meet in order to be ruled into the extension of a non-racialized adjective G can also shift depending on what comparison class is made salient by the context. For instance, consider how each instance of “smart” in the sentences below contributes to the truth-conditions of each sentence in a different way:

Tony is very smart [for a 4th-grader].

Tony is very smart [for a French bulldog].

Tony is very smart [for a theoretical physicist].

Here it seems that the standard that Tony needs to meet in order to count as “very smart” shifts as a function of the contextually salient class to which Tony belongs.

We can see where this is going: in a parallel fashion, we ought to expect the standards associated with perspectival racialized terms to be sensitive to the contextually salient *racial* class within which the individuals described by those terms belong. Following the analysis above, our gloss of the truth conditions of “lazy” and “criminal” will be as follows:

29) a. X is lazy.

b. The degree of laziness possessed by X exceeds the relevant standard for laziness at the context.

c. $\text{laziness}(X) > s(\text{laziness})$

30) a. X is criminal.

b. The degree of criminality possessed by X is greater than the standard

degree of criminality at the context

c. $\text{criminality}(X) > s(\text{criminality})$

Now, the contribution that “lazy” or “criminal” makes to the intuitive meaning of each generic is partially determined by the perspective from which the interpreter regards the conduct of the individuals or groups being described. Perspectives thus impose different *constraints* on the system interpreting the sentence. Prior to such substitution, the value of the standard parameter is indeterminate, for the simple reason that the question of whether or not X exceeds the standard for criminality depends a great deal on what X is. Substituting “Blacks” for “X” yields the following semantic gloss for two generic sentences, “Blacks are aggressive” and “Blacks are lazy.”

31) a. Blacks are aggressive.

b. The degree of aggression possessed by Blacks is greater than the standard degree of aggression at the context

c. $\text{GEN criminality}(\text{Blacks}) > s(\text{criminality})$

32) a. Blacks are lazy.

b. The degree of laziness possessed by Blacks is greater than the standard degree of laziness at the context

c. $\text{GEN laziness}(\text{Blacks}) > s(\text{laziness})$

Now, these glosses on 31) and 32) only show what must be the case in order for the sentences to be true. They do not yet provide an explanation of why 31) and 32) are more likely to be accepted than, say, “Whites are aggressive” or “Whites are lazy.” In short, they do not yet explain how these sentences could be subject to a race-related shift in meaning standards.

Here's where metasemantic perspectives come in. On the plausible view that interpreting a given gradable in context requires the interpretive system to 1) find a contextually relevant standard corresponding to the gradable, and 2) to judge whether the kind described by the gradable exceeds that standard, the term "Blacks" in "Blacks are aggressive" makes salient to the interpretive system a *different* standard that Blacks must meet in order to be judged criminal than other social kinds, whose corresponding standards would differ.

Now, here is how such a theory could give content to the notion that meaning-standards shift in the face of race (at least with respect to gradable adjectives). Allow that individual interpreters can “shift” or “recalibrate” the value of s on the fly during online interpretation. An “upward” shift in a standard would entail that an object must satisfy a higher degree value on the scale in order to be judged to fall into that gradable's extension. Conversely, a “downward” shift in a standard entails that an object must satisfy a lower degree value on the scale in order to count as falling into the extension of the gradable. I introduce the notation (\uparrow_s) and (\downarrow_s) to denote the upward and downward shifts, respectively, that occur during racialized interpretation.

33) a. Blacks are criminal.

b. The degree of criminality possessed by Blacks is greater than the
Standard degree of criminality at a context

c. $\text{GEN criminal(Blacks)} > s(\text{criminal})$

d. *Interpretation:* $\text{GEN (Blacks)} > \downarrow_s(\text{criminal})$

34) a. Lawyers are criminal.

b. The degree of criminality possessed by lawyers is greater than the

standard degree of criminality at a context

c. GEN criminal(lawyers) > s(criminal)

d. *Interpretation*: GEN (lawyers) > \uparrow s(criminal)

The prediction is that when individuals and kinds racialized-as-Black are under discussion, the standard that must be exceeded in order for those individuals and kinds to fall into the extension of Black-specific racialized gradable adjectives will *trivially* be given a lower degree value than that required for non-Black individuals and kinds. Now, what explains *why* the standards are “shifty” in this way, and what explains *why* the interpreter assigns a given semantic value during the interpretation of a given gradable is the presence of dominant social perspectives which treat the semantic values of “Black” and “criminal” as *co-varying* in some sense.

There are interesting further semantic and empirical questions in the vicinity concerning just how to understand what is going on in an interpretive system that engages in this sort of standard calibration. I will focus on just two possibilities. First, the interpretive system might use a *scale invariant* mapping of individuals to degree values. Second, the interpretive system might use a *scale variant* mapping from individuals to degree values.

To say that an interpretation of ‘criminal’ is *scale invariant* is to say that there is only one scale for criminality that all interpreters associate with the gradable adjective. Recall that a scale for a gradable X consists of a set of degrees to which an object is X , an ordering of those degrees, and a dimension associated with X . In the case of gradables like “tall,” the associated dimension is *height*. If “criminal” is indeed scale invariant, there are two ways of explaining the fact that “Blacks are criminal” is more likely to be judged true than “Whites are criminal.” The first is that the degree value which Blacks must meet in order to count as criminal is simply lower than the degree value

required for “Whites are criminal” to be judged true. The second explanation would be that Blacks and Whites receive systematically different “mappings” to the criminality scale from the interpretive system: Blacks tend to be mapped to a *higher* end of the scale by default, and Whites tend to be mapped to a *lower* end of the scale.

Let’s now consider the possibility that such mappings are *scale-variant*. It’s plausible that different interpreters can associate certain gradable adjectives such as ‘beautiful’ and ‘tasty’ with different scales (more specifically, different dimensions.) For instance, a painting might be beautiful according to a scale for *colorfulness* or according to a scale for *austerity*, and a beer can be tasty according to a scale for *refreshingness*, *lightness*, *bitterness*, etc. In the same way, it might be the case that “criminal” selects different scales for criminality depending on features of the context.

For instance, perhaps ‘criminal’ selects *violent crime* or *drug crime* when it is predicated of Blacks, and *tax fraud* or *white collar crime* when predicated of Whites or other social kinds.

Which analysis should we prefer? While these are of course largely empirical questions which require careful further work, there seems to be at least one compelling *a priori* reason to prefer the scale-invariant option. In particular, scale-invariance seems to be able to explicate the truth conditions of *comparative sentences* of the form *A is more criminal than B* (e.g. “Murdering someone is more criminal than stealing,” “John is more criminal than Jordan,” “Blacks are more criminal than Whites”) better than scale-variant options can. The reason is this: as we have seen, scale-variant options all in some way treat the scale as “indexed” to the argument taken by the gradable. We should thus expect, for example, that an argument such as *Whites* will be associated with a different scale for criminality than will *Blacks*. Regardless of what other virtues this analysis might have, it fails to explain how the truth-conditions of comparisons between Black and white levels of criminality

could be informative:

35) *Sentence*: X is more criminal than Y.

Scale-variant analysis: X is more criminal-for-X than Y.

36) Blacks are more criminal than Whites.

“Blacks are more *criminal-for-Blacks* than Whites.”

On the other hand, this might be too hasty a dismissal of scale-variance. For instance, the result of indexing specific criminality scales to particular groups need not be trivial. Perhaps when Blacks and other non-Whites are under discussion, ‘criminal’ defaultly selects perspectives from which *violent crime* or *drug crime* are most salient when the term is predicated of Blacks, and defaultly selects perspectives from which *tax fraud* or *white collar crime* are most salient when it is predicated of Whites. This would be one way of salvaging the scale-variant option, since under this analysis, the truth-conditions of comparative generics would be informative:

37) X is more *criminal-on-scale-S* than Y is *criminal-on-scale-S*.

38) Blacks are more *criminal-on-street-crime-scale* than Whites are

criminal-on-street-crime-scale.

Regardless of whether we choose to be scale-variantists or scale-invariantists, however, it seems that we have identified two interesting and informative ways in which the truth-conditional features of “criminal” and other racialized terms can reflect race-specific shifts in meaning-standards.

7. Conclusion

In conclusion, I offer a caveat and a response to a possible objection. The caveat is simply

that I have so far said nothing about how perspectives might play a role in the semantics of non-gradable expressions, including racialized count nouns like “terrorist,” “immigrant,” and “thug.” This will be a central goal of Chapters 3 and 4.

Now for the objection. Imagine a response like the following: “This is a lot of machinery to assemble just to explain how the acceptance conditions on different sorts of racialized generics differ. In fact, there are a number of simpler, more elegant accounts available: we could explain the difference in acceptance conditions between ‘Blacks are criminal’ and ‘lawyers are criminal’ simply by appealing to the different contents of common beliefs about Blacks and lawyers, for instance. Perhaps due to their priors, individuals have a higher credence that an arbitrary Black person is criminal than that an arbitrary lawyer is (Silva 2018), or there are ‘closer’ or ‘richer’ mental associations between Blacks and criminality than between lawyers and criminality (Khoo 2017), or perhaps Blacks are perceived as more prone to criminality because they are an essentialized kind (Leslie 2017, my Chapter 5). Why do we need all this (meta)semantic theorizing?”

A weak response would be to chalk this tension up to meta-theoretical taste. Those interested in epistemic or psychological mechanisms, for instance, will tend to find them for the phenomena they want to model or explain, and as long as those accounts are not incompatible with my own, we should welcome another level of description for a complex phenomenon. But the real thrust of the objection is-- what does my semantic approach to different generic acceptance conditions explain that these simpler, better known approaches do not?

Here is what these simpler approaches are silent on: why are *these* lexical items paired with *these* relevant beliefs, or credences, or stereotypes, or whatever? Why is it that these same lexical items-- “criminal,” “thug,” “terrorist,” “lazy,” “aggressive”-- seem to recruit different mental

representations depending on how things stand with the racial context? In other words, what mechanism connects an interpreter's knowledge of the lexicon with the mental representations that are the constituents of stereotypes, relevant beliefs, etc.? This is the theoretical lacuna that my account responds to.

Yet my account does have a set of testable hypotheses about these mechanisms. I have argued that for socially perspectival vocabulary, this "connecting mechanism" consists of three things: i) the logical form of linguistic expressions, ii) patterned uses of these expressions within concrete social practices, and iii) the deliverances of dominant social perspectives within those practices which determine the semantic values of socially perspectival vocabulary. This opens up space for an account of the properly *linguistic and social* mechanisms (as distinct from narrowly epistemic or psychological mechanisms) we need to make sense of the idea that racialized discourse is characterized by race-specific shifts in meaning-standards.³⁴

³⁴ In fact, we can even turn the tables on this objection. Often simpler approaches to the sorts of linguistic processing data I have discussed in this chapter insists that "background knowledge" plays a large role in determining how the interpretive system functions, and then leaves the notion of background knowledge almost entirely untheorized (e.g. Prinz 2002). The fact that I have advanced a hypothesis upon which empirical data can bear (Chapters 4-6) makes it a worthy competitor to simpler views.

Chapter 3. When Code Words Aren't Coded

Abstract

Chapter 2 showed that a theory of racialized discourse needs to be especially attentive to the fact that the social function of some linguistic items can tell us a good deal about their meaning. The current chapter extends this point. I argue that the observed racial norms within specific discourse communities can constrain the meanings of racialized terms. These norms are thus one source of what I have called race-related shifts in meaning-standards.

In defending this claim, I intervene into a contemporary debate about racial communication in political speech. The problem is this: politicians used to have to rely on “code words” to get their racial messages across. Yet recent years have demonstrated that politicians often get away with quite explicit forms of racism in mainstream political discourse. It doesn't seem that much of this speech is “coded” anymore. On the other hand, Americans in general do seem to think that racism is bad, and are hesitant to endorse attitudes that they believe to be racist. How can we reconcile the effectiveness of explicitly racialized speech with Americans' reticence to be racist?

I address this question in two steps. Negatively, I argue that the “standard framing” of racially coded speech, presupposed by much extant work in philosophy of language, is insufficient to account for the normative structure of more recent (and more explicit) racial appeals. In essence, the standard framing fails to appreciate the fact that many interpreters are incentivized to defect from more “standard” anti-racist norms in favor of the racial norms extant in their own discourse communities. More positively, the chapter diagnoses some of the mechanisms that underlie such defection. I provide an analysis of an ascendant type of explicitly racial appeal called candid racial communication. I then argue that the linguistic vehicles of candid racial communication are neither

true code words, nor slurs, but what I call *racialized terms*. I proceed to argue that we should see the normative structure of candid racial communication as having important upshots for a theory of what racialized terms mean.

“[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

0. Introduction

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 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 racial 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 now ascendant. 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 the chapter is to show that the standard framing gives unsatisfying accounts of both the *normative structure* of these openly race-specific appeals and the *meaning* of their linguistic vehicles. We will also see that the standard framing is faced by substantial empirical and conceptual problems quite independently of this explanatory limitation.

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.” Rather, they belong to a “not-so-coded” lexical category 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 this form of communication open flaunts 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 content. After showing that this hypothesis is plausible in the light of evidence

from recent philosophy of language and linguistics, I suggest that we should understand racialized terms as instances of “community-specific speech” rather than as occupying a position on the traditional “slur/code word” spectrum.

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 history and 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 widespread norms and beliefs concerning race relations. 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 emerges 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 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. Flaunting 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 they are accused of violating the norm of racial equality. Similarly, they allow audiences--even those strongly committed to the norm-- to accept the implicit racialized message without thinking that they *themselves* are violating the norm.

To see this dynamic at work, consider a comment made by then-US Representative (and current Speaker of the House) 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 males 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 a media 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 of implicit racial appeals, the central function of code words is to “open up space for deniable norm violations” (Khoo 2017) so that interlocutors can traffick 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. Rather, 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. 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 C can truthfully claim that C does not refer to any particular racial group. And on the audience side, the hearer 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 roughly, and precludes us from marking relevant distinctions between interestingly 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.)³⁵

According to the simple theory, while both the term “African-American” and the term “inner city” trigger stereotypes about Blacks, only “inner city” offers a seemingly non-racial interpretive option. Since subjects implicitly treat basing policy preferences on race-specific grounds as violations of the norm of racial equality, they only form policy preferences on race-specific grounds when they are *unaware* that they are doing so. The available non-racial reading of “inner city” explains why the subjects’ racial resentments can be effective in influencing their policy preferences. The term primes them to implicitly rely on anti-Black stereotypes even though they believe explicitly that their judgments are based on their beliefs about geographical locations rather than racialized groups.

2. Limitations of the Standard Framing

The standard framing is thus organized around four assumptions.

³⁵ 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.

A. Mendelberg's Assumptions

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. Consequently, these approaches accept two further assumptions:

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 there may be interesting differences between different types of racialized speech that should be marked by a linguistic theory.

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 of 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 clearest 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 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.³⁶

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

³⁶ 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.

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.³⁷ 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.³⁸ 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

³⁷ The qualifier “problematic” draws attention to the fact that there may be acceptable forms of race-consciousness, perhaps rooted in reparative justice or diversity considerations.

³⁸ 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)³⁹

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

³⁹ 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. For instance, does saying “Blacks are disproportionately represented on the welfare rolls” constitute a racial appeal? 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.

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 belong 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 performances that will be taken to count as norm violations.⁴⁰ 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) 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.)

⁴⁰ Of course, it 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 yet are not always seen as norm violations, and are not rejected as such.

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)

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. 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.⁴¹ 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

⁴¹ 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. 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 Ann McKinney for pointing this out.

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.”⁴²

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. That 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.”

⁴² 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 Communication⁴³

#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:

⁴³ All emphases in the quotations are mine.

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)

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,⁴⁴ 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

⁴⁴ 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.

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 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 *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 consolidates power through social and political gambits on the part of “minorities” and “immigrants” designed to take status, wealth, and power away from that “real” America.

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 and resources) and those don’t (and thus are not). Psychological evidence 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.) In many cases, the increased visibility and representation of non-White demographics are seen as 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, White Nationalist, 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.”⁴⁵ 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

⁴⁵ 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).

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 chapter, 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 semantics of its linguistic vehicles.

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 chapter we turn to the question of what upshots this has for linguistic theorizing about the particular expressions involved in candid 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 responding to these code words indicates that the semantic link between these terms and racialized meanings is rather indirect. It thus does not appear that we need resources stronger than Khoo's simple theory in figuring out how they work. I thus propose to grant that what we might call "true code words" trigger post-semantic patterns of inference or association which gets across some racialized content in the way Khoo outlines.

Yet we have seen that in contexts where norms against race-consciousness are not operative, more direct semantic relationships between racialized terms and race-specific meanings are possible. Our examples of candid racial communication indicates that there are terms which i) are *directly predicated* of particular racialized groups in a way their audiences are familiar with 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 from a taxonomic point of view. 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 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 *lexical entries* for them. If we can make good on this hypothesis, we have identified a class of expressions that put pressure on Khoo’s assertion that the simple theory “may be all we need to understand coded speech” (Khoo 2017: 32).⁴⁶

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

⁴⁶ 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,” “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.

4.1. Referential Restriction

The term “referential restriction” is borrowed from Diaz Legaspe (2018). According to one orthodoxy in the literature on slurs, slurring terms generally have “neutral counterparts” which provide the thin referential content of the slur. So, for instance, the referential content of “mick” is often taken to be identical to that of the expression “Irish-American.” Against this orthodoxy, Diaz Legaspe argues that many slurring terms either do not have neutral counterparts at all, or that these neutral counterparts do not supply the referential content of the slur. 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, the following expression seems felicitous:

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

But if the thin referential content of “micks” and “Irish-Americans” is the same, this sentence should strike us as a contradiction. The fact that it doesn’t indicates that we interpret the use of “micks” as referring to some *subset* of Irish-Americans. This is a referentially restricted use of the slur.

I am not suggesting that we understood racialized terms as slurs *per se*, and indeed I will soon suggest that we need to think of racial communication in terms of linguistic categories beyond the traditional slurs/code words dichotomy. 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 to something like *Muslim perpetrator of ideologically motivated violence*. Consider, for instance, the fact that right-wing violent White extremists are rarely labelled “terrorists,” despite their

being no less ideologically motivated than their Muslim counterparts. To this extent, it seems that racialized terms have genuinely referentially restricted uses. 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) Johann is an illegal immigrant who has committed crimes, but he is not a criminal alien.

2) and 3) indicate that the uses of “terrorist” and “criminal alien” pick out some subset of the individuals in the “neutral” extension of these terms, and that Dylann Roof and Johann do not meet the requirements for membership in that subset. Of course, 2) and 3) are bound to strike some interpreters as infelicitous. (How could Dylann Roof *not* be a terrorist? How could any criminal illegal immigrant *not* be a criminal alien?) It may be possible to see these sentences as either felicitous or infelicitous, depending on our interpretation of “terrorist” and “criminal alien,” and our general perspective on the social world (and as we will see, that would largely be the point!) Yet what matters for my purposes is i) whether it is *possible* to read these sentences as felicitous, and ii) whether those who are particularly susceptible to candid racial communication are likely to think that they are felicitous.

I think that both of these assumptions are plausible. 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)

The referent 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.

Note that it may not be very clear just which group a referentially restricted use of a term picks out in context. "Immigrant" might pick out every *non-American* resident of foreign origin, which would explain the fact that Saul herself doesn't seem to fall into the extension of the term. Or, as Saul herself suspects, perhaps "immigrant" is being used in a way that picks out groups who are specifically racialized, particularly those of Middle Eastern, Asian, or African origin.⁴⁷ On the assumption that racial presentation partially explains why Saul's interlocutors restrict in the way they do, what about the meaning of "immigrant" should we conclude from this?⁴⁸ What are the mechanisms that drive this race-related shift in meaning-standards? In the remainder of this chapter and in Chapter 4, we survey some options.

⁴⁷ To an American ear, these uses of "immigrant" are less likely to seem like cases of referential restriction. Unlike 2) and 3) above, sentences like "Jennifer Saul was born in the United States and voluntarily took up permanent residence in the United Kingdom, but she is not an immigrant" seem infelicitous. But the referential slipperiness of "immigrant" in the sociolinguistic and political climate of the UK make it plausible that at least some common interpretations of "immigrant" make the sentence seem felicitous. The semantic shiftiness of this term in particular makes it an example of what Saul calls a "protean dog-whistle" (Saul 2018).

⁴⁸ As we will see in Chapter 4, one option might be to see Saul and her interlocutor as agreeing on the basic facts of the matter--namely that Saul is from America and is now a permanent resident of the UK-- while metalinguistically negotiating over where the "threshold" for *true* immigrant status should be set for the purposes of the conversation (Plunkett and Sundell 2013.) Saul and her interlocutors can thus be seen as *advocating* for a certain way of using "immigrant" in the context which reflects their background normative and evaluative commitments. Since Plunkett and Sundell are neutral on the mechanisms that drive metalinguistic negotiations, I take this proposal to be compatible with the view of referential content I advance here.

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,⁴⁹ 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)

⁴⁹ While not explored here, 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), Tirell (2012), and Maynard and Benesch (2016).

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. 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 referential content

I have argued that the structure of the racial norms within a discourse community can affect the referential content 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.

To this extent, we should understand referentially modified racialized terms less on the model of slurs and taxonomic terms, and more on the model of various forms of "community-specific speech" (Hebert and Kukla 2016.)⁵⁰ On Hebert and Kukla's (2016) use, community-specific speech, including jargon, slang, and inside jokes, is speech that is "used by and

⁵⁰ 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.

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.⁵¹

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.)

⁵¹ 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 anti-referentialist) semantic framework (Muhlebach 2019.)

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 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 (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:

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

4) 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 4) 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 5) by use of 6):

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

6) 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 6) does not directly target various other components of 5). 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:

7) Wait, it didn't rain today!

8) I didn't know you had a bike!

9) 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:

10) The main problem that faces the national budget is welfare.

11) No it isn't!

A speaker of 11) targets the at-issue content of 10) 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.

12) I rode my bike to the park in the rain. # It wasn't raining.

13) 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, 14) and 15)

should sound odd, since they explicitly cancel not-at-issue contents associated with one or more terms in the discourse:

14) These tax breaks help **inner city** families. Most of them are White.

15) 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 14) and 15). 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 (or 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: it seems that 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.

My core proposal-- that the observed racial norms at a context constrain the referential content of a racialized term in a way that is highly conventional-- 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 on the one hand and social-scientific explanation on the other.

Let me explain. Ken Taylor (2007) has introduced the term “modificationally 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.⁵² 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.⁵³ In effect, modificationally 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

⁵² 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.

⁵³ What we might call “radical contextualists” (e.g. Travis 1985) deny that the problem is one for metaphysics rather than a theory of meaning. Taylor’s invocation of modificationally neutrality is explicitly intended to show that we need not follow the radical contextualist on this point.

seamlessly honor different sorts of modifications for the property expressed by “red” in a way that 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 systematically 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 entirely 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 intuitive meaning 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

Now, I do not intend for this to be the final word about the meaning of racialized terms. The point of the moment is simply that my proposal about the referential content of these terms is a promising way of explaining how one sort of race-related shift in meaning-standards occurs. That proposal has the added benefit of representing a dialectical middle ground between Khoo’s minimalistic proposal and Stanley’s implausible proposal about racialized meaning. In Chapter 4, we will investigate some other possible sources of race-related shifts in meaning-standards. There we will present a much more ambitious picture of how racial features of the context and linguistic meaning interact. In particular, we can show that such shifts occur at the level of *standing meanings*, given certain plausible metasemantic assumptions.

Chapter 4. The (Meta)Semantics of Racialized Terms

Abstract

Chapter 3 argued that we need resources beyond the standard framing to account for the meaning and use of racialized terms. There we were concerned to show that race-specific shifts in meaning-standards might occur at the level of referential content. In this chapter, we make a much stronger claim: race-specific shifts in meaning-standards might occur at the level of *lexical meaning* or *standing meaning*. The purpose of this chapter is to show what such an account might look like.

Minimally, a theory of racialized terms ought to explain i) why it is that racialized terms seem to propose different criteria for membership in the extension of the term in a way that depends on race, and ii) why the race-specific content of these terms seems cancellable in at least some contexts. My goal is to meet these requirements while assembling evidence for the idea that the variable lexical meaning of racialized terms is responsible for these two behaviors.

As such, this chapter presents my metasemantic assumptions in detail. I argue that racialized terms-- and perhaps other bits of vocabulary as well-- have three interesting metasemantic properties: i) they are socially perspectival, ii) they are context-sensitive, and iii) their meanings are underdetermined. After demonstrating the plausibility of these assumptions, I show how they might be brought to bear on a first-order semantics for at least some racialized terms. That semantics reviews evidence that terms like “thug,” “criminal alien,” “terrorist,” and “immigrant” are *inherent generics*, whose lexical structure includes contextual parameters whose semantic values can be set by (race-specific) features of the context. Of particular note is the fact that this metasemantics and first-order semantics nicely integrates with some of the psychological data on racialized discourse, which will be the focus of Chapters 5 and 6.

[The classical Fregean] view conceives of language as a railway line designed by a corps of volunteers (the advance party sent out at the “setting up” of the language) and the process of employing it consists in sitting in the observation car and remarking on the scenery passed en route. I advocate thinking instead of those comical locomotives in children’s literature which unroll their tracks before them as they move through a terrain. In the short run, the path of the train responds to the hills and gullies of the landscape virtually as well as the better designed line and the view from the observation car may seem practically identical. In the words of the old song, “she gets there just the same,” partially because the passengers are along only for the ride and have no clear conception of a long-term destination anyhow.

--Mark Wilson, “Predicate Meets Property,” 1982

0. Introduction

Khoo’s approach to the semantics of code words holds that code words are effective in political communication just in case they trigger patterns of inference or association that produce race-specific contents. He also suggests that the simple theory “may be all we need to understand the semantics of code words.” Unfortunately, the simple theory is not so much a “theory” as it is an invitation for psychologists to do more research. The simple theory certainly describes how code words behave, but it doesn’t explain why they behave that way. Moreover, the simple theory tells us virtually nothing about how, say, the term “inner city” gets across its racial message in a way that is relevantly similar or different to the way that the Confederate flag conveys the central tenets of White Supremacy.⁵⁴ The simple theory is so broad, and so agnostic about the mechanisms that drive interpretation, that it seems not only unfalsifiable, but also unable to explain at any deep level why *language in particular* can be an effective means of transmitting racist beliefs at all.⁵⁵

⁵⁴ To be fair, Khoo believes this lack of differentiation between linguistic and non-linguistic “communication” to be a virtue of his theory, insofar as it might reveal how code words and “social meanings” interact. One philosopher’s *reductio* is another’s reason to believe.

⁵⁵ Again, to be fair, Khoo’s theory receives some negative support. Khoo argues that extant approaches to code words, particularly Mendelberg (2001) and Stanley (2015), cannot account for the behavior of code words as long as they hold that such terms have well-defined racialized and non-racialized meanings. In a series of compelling arguments, Khoo shows that the term “inner city” is neither ambiguous between *urban core* and *lazy Blacks*, nor can *lazy Blacks* be considered the “not-at-issue content” of “inner city.” I propose to grant Khoo those arguments, since my theory of racialized terms turns on none of the claims he attacks.

I suggest that providing a more satisfying account requires us to specify the role of *semantic content* in determining what is communicated. The challenge of the present chapter will be to provide some general characterizations of this role across a wide variety of linguistic expressions while also specifying how the semantic content of racialized terms in particular allows them to communicate race-specific messages. This will require us to toggle back and forth from claims about linguistic meaning in general to claims about the meaning of racialized terms in particular.

Consequently, the chapter is divided into two major parts. The goal of Part 1 (Sections 1-3) is to set out some criteria of adequacy on an account of the meaning of racialized terms, and to review some theoretical options in meeting those goals. The goal of Part 2 (Sections 4-6) is to provide a unified semantic and metasemantic account of the meaning of racialized terms. This account not only meets the criteria set out in Part 1, but also affords us powerful expressive possibilities for understanding the mechanisms behind race-specific shifts in meaning-standards.

Section 1 provides a more detailed argument for thinking that racialized terms comprise a distinct semantic class. I show that racialized terms can not only not be reduced to the class of code words, but also that they cannot be thought of as either slurs or purely taxonomic terms. Section 2 sets out some plausible criteria of adequacy on a semantics for racialized terms. The semantic account should explain i) why racialized terms propose different semantic *thresholds* for different racialized groups, and ii) why the racial content of racialized terms is *defeasible*. Section 3 sets out a few different strategies for meeting these criteria. I introduce a working conception of the distinction between “semantics” and “pragmatics,” and suggest that contrary to orthodoxy, the race-specific content of racialized terms should be accounted for by a *semantic* theory of those terms.

Beginning Part 2, Section 4 sketches my approach to the metasemantics of racialized terms-- what facts make it the case that racialized terms mean what they do in context. My core assumptions are that the meanings of racialized terms (among others) are *underdetermined* as well as *context-sensitive*. After drawing on relevant evidence for this claim, I reintroduce the notion of a social perspective (last discussed in Chapter 2) in order to show how these underdetermined and context-sensitive meanings become more determinate in context. Section 5 moves from the metasemantics of racialized terms to their semantics. I take it that many first-order accounts of the semantics of racialized terms are compatible with the metasemantic assumptions I put forward in Section 4. Focusing on a subset of racialized terms (count nouns including “thug,” “terrorist,” “immigrant,” and the like), I show that it is useful to think of these racialized terms as *inherent generics*. As we will see in Sections 5 and 6, this treatment makes clear that the sorts of contextual elements social perspectives might manipulate are included in the semantics of racialized terms. This is thus a vindication of the view that racialized terms contain race-specific content. Finally, Section 7 shows how this combined metasemantic and semantic account meets the two criteria of adequacy introduced in Section 2: the “threshold problem,” and the “defeasibility problem.”

PART I

1. What are racialized terms (again)?

Philosophers of language have written a fair amount about slurs and deeply derogatory terms. They also have a bit to say about the semantic properties of taxonomic terms for human races such as “Black,” “White,” etc. As we saw, they have also begun to address the linguistic properties of

code words and dog-whistles. To glance at the extant literature, one might believe that these are the only three lexical categories relevant to the intersection of race with linguistic meaning.

Yet racialized terms are not easily assimilated to any of these classes. They are not slurs, since they lack strongly derogatory content. Unlike slurs, they are generally accepted in political discourse. These terms are not code words, since i) they do not “code” their racial content, and ii) their effectiveness does not depend on their being believed to be “deniable norm violations.” To this extent, they function more like taxonomic terms. Yet they are not *merely* taxonomic terms, since they pick out kinds which are racialized in some circumstances, but not others.

The class of is thus negatively defined by *not* quite being slurs, code words/dogwhistles, or taxonomic terms. Let’s briefly consider some reasons to think that racialized terms belong in their own semantic class. Since Chapter 3 has already made the case for thinking that racialized terms are not code words, I just touch on slurs and taxonomic terms here.

1.1 A negative definition of racialized terms

(Non-reappropriated) slurs are terms which strongly derogate groups either on the basis of group membership, or on the basis of some property that is deemed to be central to the group. Philosophers and linguists generally agree that an account of slurs must do two things. First, it must provide an account of what the *referential content* of slurs are, or explain why they lack such referential content. Second, it must provide an account of the linguistic basis of the *derogatory force* of slurs.

An interesting feature of slurs is that they retain their derogatory force even when they do not have the force of *assertions*. Rather, this derogatory force intuitively “projects” even when these

terms are embedded in non-assertional contexts. So, assuming that “mick”⁵⁶ is genuinely derogatory, an “aura” of derogation remains even when speakers do not assert that anyone in particular is a Mick:

- 1) *Assertion*: Patrick is a mick.
- 2) *Interrogative*: Is Patrick a mick?
- 3) *Conditional*: If Patrick is a mick, so is Seamus.
- 4) *Negation*: Patrick is not a mick.

One possible explanation for the idea that 1)-4) can each seem derogatory in some sense is that speakers seem to inherit a commitment to regarding Irish-Americans in a highly derogatory way simply by virtue of uttering the term. (Interestingly, this is even true of 4), which explicitly denies that Patrick belongs into the extension of “Mick.” Plausibly, this is because even direct negation does not challenge the legitimacy of a certain way of thinking or speaking about Irish-Americans.) This “projection” of derogatory content often taken as evidence that derogatory force is linguistically encoded in slurs, and that such content could float free of speaker intentions.

Closely related to this behavior is the fact that the derogatory force and referential content of slurs are *non-cancellable*. That is, once the slur is used in the context of an assertion, the speaker is committed to an interpretation of the slur’s referent, and perhaps even a certain way of thinking or feeling about that referent. For instance, without further linguistic moves, the following sentences seem infelicitous:

- 5) # Patrick is a mick, but he’s not Irish-American.

⁵⁶ Since the point can be made without mentioning vicious slurs for subordinated racial and ethnic groups, I here assume for the sake of argument that “mick” is a genuinely derogatory slur for Irish-Americans, even if there are far more objectionable slurs out there. Readers are invited to think of “more derogatory” slurs if the point is not clear.

6) # Patrick is a mick for reasons that have nothing to do with his ethnicity.

Even derogatory racialized terms do not share either of these features of slurs. First, even when dehumanizing terms like “savage” are applied to racialized groups, the derogatory force of “hood rat” does not seem to project outside of assertional contexts:

7) *Assertion*: Tayla is a savage.

8) *Interrogative*: Is Tayla a savage?

Similarly, whatever racialized content might be associated with “savage” can be felicitously cancelled:

9) Tayla is a savage. She’s from a White lower-class background.

10) Tayla is a savage regardless of what race she is.

Finally, this free cancellability of racialized terms also explains why they are distinct from taxonomic terms. While it is impossible for Patrick to fall into the extension of “Irish-American” or “mick” without his membership being *explained* by his racial membership, it is clearly possible for Tayla to be a “savage” quite independently of her racial membership (although this may be different in different discourse communities.) This seems to suggest that if “savage” contains racialized content, that content is not part of its “core meaning.”

To sum up these similarities and differences:

Term Type	Derogatory force “projects”?	Directly predicated of racialized groups?	Defeasible racial content?
Racialized terms	NO	YES	YES
Taxonomic racial terms	NO	YES	NO
True Code Words	NO	NO	YES

Racial Slurs	YES	YES	NO
--------------	-----	-----	----

Similarities and Differences among types of Racialized Discourse

Figure 3

2. What Should a Semantics for Racialized Terms Explain?

A semantics for racialized terms should explain why they display their characteristic linguistic behavior. We can thus extract two desiderata for a semantics for racialized terms. Call these the *Threshold Problem* and the *Defeasibility Problem*.

Threshold Problem: What about the meaning of racialized terms (if anything) allows them to propose different thresholds for differently racialized agents?

Defeasibility Problem: What about the meaning of racialized terms (if anything) allows their racialized “content” to be felicitously cancelled?

Assembling an account that answers these three questions will give us a more positive conception of how racialized terms work in linguistic communication. Let’s briefly look at each problem in more detail.

2.1 *Threshold Problem*

In Chapter 1, we noted that some terms are often applied to a racialized group A and not to a racialized group B, even if the conduct of A’s and B’s are type-identical. For instance, Muslims who engage in mass murder are described as “terrorists,” often regardless of their motivations, while Whites who commit mass murders in schools, for example, are rarely so described. This is evidence that different racialized groups face different *thresholds* for the application of certain terms. A theory of racialized terms ought to explain why this is the case.

Here there are two basic options. The first option is to see these differential thresholds purely as a matter of the *use* of certain terms. On this approach, terms like “terrorist” have a fixed, race-neutral meaning, but are selectively applied to Muslims in particular. Solving the threshold problem would require us to explain why the term has come to be used in this selective way. Understanding this pattern of use might require us to appeal to the racial biases of language users. But we wouldn’t say that “terrorist” is differentially applied to Muslims viz. Whites by virtue of having the property *Muslim* as “part of its meaning.”

The second option is to see the differential thresholds as having something to do with the *meaning* of racialized terms. On this approach, differential thresholds across racial groups are *not* merely features of the *use* of these terms. Rather, the existence of the differential threshold is explained (at least in part) by the *meaning* of the term in question. So, the fact that “terrorist” is selectively applied to Muslims and rarely applied to Whites is partially explained by the fact that “terrorist” just *means* Muslim (in a sense to be made clear) within a discourse community.

As we will see, most philosophers assume that the first approach is best. Yet contrary to orthodoxy, I will argue that the threshold problem can be solved by appealing to the meaning of racialized terms rather than simply their use.

2.3 Defeasibility Problem

We’ve seen that code words and racialized terms have defeasible racial messages. This means that whatever racialized information they carry can be “cancelled.” So, for instance, imagine that a use of 11) leads an interpreter to believe that Robert is Black, and that his race partially explains why he is a thug:

11) Robert is a thug.

The speaker of 11) can easily cancel this information with further linguistic moves.

12) Robert is a thug. He's White.

13) Robert is a thug for reasons having nothing to do with his race.

A theory of racialized terms needs to explain why race-specific information is cancellable in this way.

As before, we can explain this defeasibility by reference to facts about the *meaning* of racialized terms, or by reference to facts about their *use*. As we will see below, philosophers generally take the latter strategy to be best.

3. Semantics and Pragmatics

Philosophers and linguists often draw a distinction between two basic types of linguistic meaning. The first type of meaning encompasses the information which is encoded in the linguistic items of a particular natural language. This linguistically encoded information plays a number of different roles. Candidate roles may include explaining why a term like “dog” refers to all and only *dogs* (and not cats or lizards) and why a sentence like “I’m not getting out of bed today” expresses the proposition *Patrick isn't getting out of bed on whatever day is today*. The second type of meaning encompasses the sort of information which is otherwise conveyed by the use of linguistic items in context. This sort of information is not linguistically encoded in these items, but is communicated by virtue of the interaction of linguistic items with non-linguistic features of the context.

Philosophers tend to assume that the job of *semantics* is to specify how linguistic items mean things by virtue of the first, linguistically encoded, conventional sort of meaning. The job of *pragmatics* is to explain how linguistic items mean things by virtue of the second, non-encoded, context-specific, sort of meaning. Yet while there is widespread agreement *that* these two senses of meaning exist, there is much less agreement on just where “semantic” meaning ends and “pragmatic” meaning begins. There is not even widespread agreement on what the meta-theoretical terms “semantic” and “pragmatic” mean, nor is there agreement on whether successful linguistic communication is primarily explained by “semantic” or “pragmatic” meaning. As Jeff King and Jason Stanley (2005) point out in a useful discussion,

“...[A] complicating factor in these debates is the lack of a clear and accepted criterion among philosophers of language and linguists for what counts as semantic versus what counts as pragmatic. That is, among philosophers of language, there is no stable agreement on the semantics/pragmatics distinction. Furthermore, even among those who agree on terminology, there is disagreement about the scope of semantic content. That is, it is a subject of much current debate how much of what is intuitively communicated is constituted by semantic content. (King and Stanley 2005)

My goal isn't to clean up this morass, nor even to say anything original about the semantics-pragmatics distinction. Rather, it is to borrow King and Stanley's own plausible interpretation of the relationship between semantics and pragmatics, and ultimately to show that this interpretation gives us good reason to think that the race-specific meanings of racialized terms belong to the semantic profile of these terms.

What sort of linguistic meaning is “semantic” and which is “pragmatic?” Following King and Stanley's (2005) discussion, let's revisit the notions of “standing meaning” and “referential content” (see Chapter 3, Section 5.) All the expressions *E* of a language have what Richard Heck (2001) calls a “standing meaning.” The standing meaning of *E* is determined by the conventional rules of the

language. It is what a speaker grasps when she uses E correctly. Many terms have standing meanings which allow them to pick out the same thing in all contexts. The standing meaning of names such as “Aristotle,” for instance, allows them to denote the same individual in all possible worlds.⁵⁷ Yet at the other end of the spectrum, terms like “I,” “today,” “tomorrow,” “that,” and “here” have standing meanings which allow them to pick out lots of different things in a way that depends upon the context in which they are used. For instance, the “I” in “I am hungry” picks out *Patrick* when I utter the sentence, but “I” picks out *Keisha* when she utters it.

So the standing meaning of expressions like “Aristotle” allows them to pick out the same object or property in all contexts, and the standing meaning of expressions like “I” allows them to pick out different objects or properties depending on the context. Call whatever object or property E picks out at a context the “referential content” of E. Putting things in terms of “standing meaning” and “referential content” considerably clarifies the original question of which types of meaning are “semantic” and “pragmatic.” Minimally, *semantics* must provide an interpretation of the standing meaning of expressions, and say how these standing meanings allow these expressions to secure their referential contents. To this extent, we can equate “referential content” with *semantic content*.

Now, expressions frequently communicate much *more* than their referential contents. Keisha walks into a bar. Rather than serving her, the bartender asks “are you old enough to drink?” Rolling her eyes, Keisha responds, “I’m 32 years old.” From the point of view of semantics, Keisha’s response expresses the proposition *Keisha is 32 years old*. Yet from the point of view of pragmatics,

⁵⁷ As Kripke’s (1980) famous discussion shows, names have particularly rigid standing meanings. “Aristotle” picks out the same individual [Aristotle] regardless of whether [Aristotle] satisfies any of the descriptions (“Greek philosopher,” “student of Plato,” etc.) he actually satisfies in our world.

Keisha's response also expresses the proposition *Keisha is old enough to drink*. This latter proposition is no part of the semantic content of the expression "I'm 32 years old." Yet Keisha manages to communicate this proposition by getting the bartender to infer that she is old enough to drink from the proposition that she is 32 years old. In order for the bartender to infer *Keisha is old enough to drink* from *Keisha is 32 years old*, he needs to have knowledge of lots of things besides the semantically encoded linguistic meaning of "I'm 32 years old." Plausibly, he at least needs to know what Keisha *intends* to communicate by means of her utterance, and what the legal drinking age is. This mutual knowledge of extra-linguistic facts is what enables Keisha's utterance to communicate the second proposition, which goes beyond the standing meaning of the sentence she uses.

Thus the "semantic" meaning of E is given by E's standing meaning and its referential(/semantic) content. The "pragmatic" meaning of E includes whatever E communicates over and above its standing meaning and referential content. This gives us a working conception of the semantics-pragmatics boundary. Of course, this working conception doesn't answer the question of *how much* of what E communicates is attributable to its standing meaning and semantic content, and how much is attributable to post-semantic or "pragmatic" meaning. Those with a rather inflationary conception of semantics could hold that a great deal of what is communicated via E is linguistically encoded, and those with an inflationary conception of pragmatics could hold that very little of what is communicated via E is linguistically encoded. It is also possible to negotiate the boundaries of the concepts "standing meaning" and "referential content." These are "metasemantic" issues that we will explore in the next section. Yet for now we can now turn to the question of whether threshold variance and defeasibility is explained by the *semantics* or *pragmatics* of racialized terms.

There are good *prima facie* reasons to adopt a pragmatic approach to racialized meaning. As we saw in Chapter 3, terms which can function as code words/racialized terms intuitively do not have racialized groups or properties as referential contents in all contexts. The standing meaning of terms such as “inner city,” “terrorist,” and “criminal,” for instance, intuitively select non-racialized referential contexts such as *urban core*, *violent extremist*, and *law-breaker* at many contexts. Assuming, as most philosophers do, that the standing meaning of an expression is relatively invariant across contexts, this is strong evidence that standing meaning is not what explains these terms’ ability to carry race-specific messages. This of course is the approach taken by Khoo’s simple theory: the standing meaning of code words selects non-racialized referential contents, and any race-specific propositions expressed by those terms are due to post-semantic inferences. Part of the point of Chapter 3 was to put some pressure on that view.

Moreover, as mentioned above, treating racialized meaning as “pragmatic” provides a straightforward solution to the threshold problem and the defeasibility problem. Racialized terms propose different thresholds for different groups because the contexts in which people complain about “thugs,” “immigrants,” and “terrorists” are arranged in a way that leads interpreters to infer that *Blacks*, *Latinos*, and *Muslims* are under discussion. The challenge for a pragmatic theory would then be to specify how contexts must be arranged in a way that prompts these inferences. Likewise, the fact that such inferences are “detachable” from the standing meaning of these terms provides a ready explanation for why those inferences can be cancelled in context.

However, despite its intuitive plausibility, actually worked-out pragmatic approaches to racialized terms are in strikingly short supply. There has been no systematic attempt, for instance, to treat code words/racialized terms as carrying racialized *implicatures*, although the utility of such an

approach seems obvious.⁵⁸ To be sure, perhaps a “pragmatic” account along these lines could indeed account for the threshold problem and the defeasibility problem. My goal here is simply to show that “pragmatic” approaches are not the only game in town in making sense of the meaning of racialized terms.

PART II

4. Metasemantics: Context-Sensitivity, Meaning Underdetermination, Social Perspective

Settling on a working interpretation of the boundary between semantics and pragmatics does not settle all the questions we might have about *how much* of the meaning of E at a context is “semantic” and how much is “pragmatic.” We may still want to know how much information is encoded in the standing meaning of E, and whether that standing meaning is flexible enough to explain why E takes different referential contents at different contexts. Adapted to the case of racialized terms, our question is: does a racialized term T refer to racialized groups by virtue of its semantically encoded content? Ultimately, I will answer this question in the affirmative.

Yet we need to clarify just what sort of information I am assuming to be properly “semantic.” Answering this question will give us an idea of what sort of features of linguistic meaning a semantic approach to racialized terms must account for. In this section, I will defend three theses about linguistic meaning. Thesis 1 is rather uncontroversial: the linguistic meanings of some expressions are *context-sensitive*. Thesis 2 is more controversial: the linguistic meanings of most expressions are *underdetermined*. That is, the “standing meaning” of most expressions does not select a

⁵⁸ A notable exception is Eric Swanson (forthcoming), who argues that *slurs* cue harmful ideologies by virtue of carrying “acceptability implicatures” which implicate that the ideology cued by use of the slur is acceptable. Swanson is clear that such implicatures are not *conventional* implicatures on the one hand (since that would require the entire content of the ideology to be “contained” in the slur itself) nor *conversational* implicatures on the other (since the ideological content of a slur cannot be easily cancelled in most contexts.) While racialized terms are different from slurs, the point is simply that it’s not obvious how to cash out the claim that racialized content is “implicated,” and that such work has largely not been done to date.

determinate referential content independently of concrete linguistic exchanges. Thesis 3 is perhaps the most controversial, yet also the most familiar: the linguistic meanings of some terms are *socially perspectival* in that dominant social perspectives in the context of use fix the referential content of perspectival terms. Considered separately, each thesis is true of some subset of expressions. Racialized terms lie at the point at which these three claims overlap: they are *context-sensitive*, *underdetermined*, and *socially perspectival*.

4.1 Context-Sensitivity

Utterances of linguistic expressions take place in contexts which have a great many non-linguistic features. These non-linguistic features often affect the meaning of linguistic expressions. As we saw, there is an obvious sense in which non-linguistic features of the context can affect the *pragmatic* meaning of expressions. Indeed, the entire phenomenon of conversational implicature is a result of the interaction of linguistically meaningful items with non-linguistic contexts. What we are interested in determining here is what the properly *semantic* role of context is. That is, we are interested in whether context plays a role in “supplementing” the standing meaning of E in a way that allows E to secure a referential content in context.

Context plays a role in determining semantic meaning of an expression E insofar as the standing meaning of E must be supplemented by non-linguistic features of the context in order for E to secure its referential content. So, names like “Aristotle” are paradigmatically *context-insensitive* terms because their context-independent standing meaning is sufficient to pick out the same object in all contexts. Their referential content thus does not vary depending on how things stand with the non-linguistic context. On the other end of the spectrum, indexicals like “I” are paradigmatically

context-sensitive terms. “I” has a constant standing meaning which must be supplemented by contextual information--in particular, who the speaker at the context is-- in order to secure referential contents in particular contexts. Moreover, because this contextual information changes from context to context-- that is, different contexts often have different speakers-- the referential content of “I” thus can switch from context to context.

We thus have in hand a ready characterization of how context can play a role in determining both the “semantic” and “pragmatic” meaning of an expression in context:

The first way context may determine what is communicated is by affecting the semantic content, via resolution of the referential content of context-sensitive elements in the sentence uttered. This roughly corresponds to what Stanley and Szabo (2000, pp. 228-9) and Perry (2001, pp. 42ff.) call the semantic role of context. The second way is that context plays a role in determining what is communicated by the linguistic act over and above its semantic content. This is the genuinely pragmatic role of context (Ibid., pp.230-1). (King and Stanley 2005)

So, context is relevant to the semantics of E insofar as it provides E with the contextual information it needs in order for its standing meaning to secure a referential content in context. Or, to put it more pithily in the terms of Szabo (2001): “The [semantic] content of an expression depends on context only insofar as the [semantic] contents of its constituents do.”

Many cases of semantic context-sensitivity, such as indexicals and demonstratives, are uncontroversial. Yet many philosophers and linguists take semantic context-sensitivity to be a fairly limited phenomenon displayed by indexicals, demonstratives, and perhaps a few other scattered expressions. However, the past several years have seen an explosion of work on semantic context-sensitivity beyond these few paradigm cases. A wide variety of texts now argue that epistemic modals like “might” (von Fintel and Gillies 2011, Silk 2016), gradable adjectives like “hot,” “cold,” and “tall” (Kennedy 2005, Glanzberg 2007), predicates of personal taste such as “tasty,”

“spicy,” and “disgusting” (Glanzberg 2007, Glanzberg ms), relational terms such as “friend,” “enemy,” and “brother” (King 2014), and even social kind predicates such as “man” and “woman” (Diaz-Leon 2016), are context-sensitive with respect to their semantics.

Unlike indexicals and demonstratives, none of these expression types are *obviously* good candidates for being context-sensitive. For some, there is a powerful intuition that the referential content of these terms is invariant once their standing meaning is fixed, and that this standing meaning is fixed independently of context (e.g. Cappellen and Lepore 2005). Consequently, showing that a seemingly context-insensitive expression E is in fact semantically context-sensitive requires some work. Arguments for the context-sensitivity of E have a common form. First, it must be demonstrated that the referential content of E shifts systematically depending on context. Second, we must rule out the possibility that shifts in E’s referential content are due to something other than elements of E’s standing meaning. That is, shifts in E’s referential content must not be attributable to the “pragmatic” role of context. One particularly powerful way of ruling out this possibility is to show that E contains elements in its syntax which can be fixed by information in the extra-linguistic context.

This is precisely the strategy that I will exploit below to argue that racialized terms are semantically context-sensitive. But first, let’s see how the strategy works in the case of one type of context-sensitive expression in particular, predicates of personal taste. The example will be useful insofar as it shows us what sorts of evidence could be relevant to the question of whether an expression is context-sensitive or not.

Predicates of personal taste (PPTs) include gustatory expressions like “tasty,” “salty,” and “disgusting” as well as aesthetic expressions like “cool,” “ugly,” and “beautiful.” There are two

sources of context-sensitivity in the standing meaning of PPTs. First, PPTs are gradable adjectives. As we saw in Chapter 2, gradable adjectives are scalar terms. To describe X as “tasty” is to locate X on a scale for gustatory quality. This gustatory quality scale is associated with “tasty” by virtue of its standing meaning. Yet this standing meaning needs supplementation-- in particular, context must determine the value of the contextual threshold for gustatory quality. “X is tasty” will be true just in case X’s degree of gustatory quality exceeds this contextually determined threshold. The standing meaning of PPTs thus contain a standard parameter whose value is fixed by the context.

This much is familiar. Yet predicates of personal taste are subject to a second source of context-sensitivity. Let’s say I have a sip of beer and exclaim:

14) Wow, that’s tasty!

You, naturally, will come to believe that the beer is tasty *to me*, according to my perspective on things. Moreover, it seems that without knowing who made the report, you don’t know whether 19) is true or false. For instance, 14) might be true from the perspective of taster 1, but false from the perspective of taster 2. This suggests that the perspectival nature of “tasty” is a proper part of its semantic profile. More generally, the standing meaning of any predicate of personal taste “X” seems to be something like *X-to-experiencer-E*.

Michael Glanzberg (2007, ms) has argued that PPTs bring two “contextual parameters” to the context of use. The first is the familiar standard parameter proper to gradable adjectives, and the second is an “experiencer” parameter which belongs only to PPTs. Glanzberg appeals to a variety of syntactic evidence in support of the claim that this experiencer parameter is seen by the semantics of PPTs. First, many predicates of personal taste naturally accommodate *to* and *for* adjuncts:

- 15) Sushi is tasty *to me*.
- 16) That joke was pretty funny *to me*.
- 17) Skiing is fun *for everyone*.

Intuitively, each adjunct phrase modifies or describes something which was already stated by the non-italicized phrase. In many cases, explicit adjunct phrases are superfluous, since the information they provide is supplied by the context. The fact that these “to/for E” adjuncts are optional and even redundant suggests that the perspective of E is already encoded in the semantics of the predicate, perhaps in the form of an argument place.⁵⁹

Moreover, it does seem that predicates of personal taste incorporate genuine *argument* places which are syntactically represented but phonetically unarticulated (in short, just the sort of things that should appear in an accurate lexical semantics for these terms.) One source of evidence here is the fact that the experiencer class associated with a given predicate of personal taste is not *iterable*. In general, arguments are not iterable, while syntactically unrepresented “adjuncts” are:

18) Iterable adjuncts: Bob read the book that fell off the shelf by the bed that was dropped off by the truck that left the warehouse that...

19) Non-iterable arguments:

- a. #) John kissed Bob Keisha Richard Bella
- b. #) Shostakovich’s music is beautiful to me to you to Bob to everyone...

⁵⁹ Sometimes adjuncts *are* required to add information that wouldn’t be available otherwise. Consider “Lobster might be tasty to you, but it’s not tasty to me,” in which the perspectival positions must be made explicit in order to draw out the nature of the disagreement over the gustatory quality of lobster.

c. #) For me for you for all of us skiing is fun.

Persuaded by these and other arguments from syntax, Glanzberg (ms) argues that predicates of personal taste incorporate an *experiencer* parameter **E** which takes as its semantic value the perspective of the agent(s) to whom something is tasty, beautiful, fun, etc. This means that predicates of personal taste bring *two* distinct contextual parameters to the context of use-- the standard parameter, *s*, and the experiencer parameter **E**. Hypothesizing these two parameters yields the following rough semantic gloss for predicates of personal taste:

20) [[tasty]] = **exceeds-gustatory-standard-S-to-experiencer E**

21) [[beautiful]] = **exceeds-aesthetic-standard-S-to-experiencer E**

Predicates of personal taste thus have a standing meaning which requires contextual supplementation in order to secure their referential content. The point of working through this example has been to show that we don't need to rely simply on our intuitions concerning whether an expression E is context-sensitive or not. Rather, we can appeal to facts about the syntax and communicative function of E to build a compelling case for the semantic context-sensitivity of E.

4.2 Meaning Underdetermination

Our discussion of linguistic meaning so far has taken for granted that “the conventional rules” of a natural language L moderately-to-strongly constrain the standing meanings of the expressions in L. This is a common methodological and empirical assumption in most analytic philosophy of language. However, there is also a philosophy of language tradition inspired in turn by generative linguistics, “ordinary language” philosophy, dynamic logic, and psycholinguistics which is

more skeptical of the notion of “standing meaning,” and likewise skeptical that the “conventional rules” of a language strongly constrain the meanings of expressions within it.

The thesis of *meaning underdetermination* is that an expression E is compatible with a variety of determinate(ish) meanings m_1, \dots, m_n and that the question of “which” meaning E has is underdetermined by E’s syntactic and semantic structure. The idea is radical insofar as it entails giving up on the idea that the rules of a language select a *privileged* determinate meaning for each expression in the language. Yet it is conservative insofar as it still posits constraints on what expressions *can* mean. According to the thesis, the meaning of E can become *more* determinate within the context of particular linguistic exchanges. Here what lends determinacy to E’s meaning is not the rules of L, but the various ways in which conversational participants negotiate the meaning of E “on the fly.”⁶⁰

Within philosophy, Peter Ludlow (2014) has given the most thorough defense of the thesis of meaning underdetermination. Ludlow argues that analytic philosophy of language has been misled by a picture of language according to which linguistic meaning is “static.” According to this static picture, language is “a widely shared common currency that agents use to communicate, with

⁶⁰ Note that the thesis of meaning underdetermination and the thesis of semantic context-sensitivity are compatible but distinct theses. Context-sensitivity says that context is semantically relevant insofar as it plays a role in fixing the value of contextual parameters within the standing meaning of a context-sensitive expression E. Meaning underdetermination says that the meaning of E is underdetermined i) regardless of whether E is context-sensitive, and ii) continues to be underdetermined even when the values of E’s contextual parameters are fixed. As Ludlow puts it,

Note that there is still underdetermination [of the meaning of “I”] once the parameter is fixed. That is, if the speaker who utters “I” is me, the parameter is set to me, but does that include my hair? Clothes? Glasses? In the game of baseball a ball can “hit me” by hitting only my shirt or hitting my hands, but I can still lose both my shirt and hands and still be me. Similarly, if I utter “now” the time interval is fixed to the time of utterance, but is that this minute? This nanosecond? Today? This geological era? Again, I take this to be underdetermined. (Ludlow 2014: 87)

individual words being the common coins of the realm. These common coins are also supposed to be more or less fixed” (Ludlow 2014: 2).

Yet this static picture seems to clash with how actual human speakers use their words in flesh-and-blood communication. Perhaps the main source of evidence for the plausibility of meaning underdetermination is the myriad negotiations, stipulations, re-interpretations, reappropriations, and contestations over linguistic items we engage in during the course of everyday discourse. Anyone who has gotten into an argument over whether a hotdog is a sandwich, or whether kombucha counts as an alcoholic beverage, whether justice really is the will of the strong over the weak, or whether Black separatists count as a “hate group” is intimately familiar with the sort of lexical maneuvers, false starts, contestations, and stipulations we use to narrow, expand, and sharpen the meanings of our terms within particular linguistic exchanges.

In a recent series of papers, David Plunkett and Tim Sundell (2013) have identified one type of linguistic contestation they call *metalinguistic negotiation*. Often we use a word X in order to say something about the world. Call this a “linguistic” use of X. Yet we also frequently use X in order to say something about what X itself *should mean*. Call this a “metalinguistic” use of X.

Consider first a disagreement involving linguistic uses of the term “dog.”

22) *Context: An animal walks into a poorly lit room where A and B are sitting.*

Richard: That’s a dog.

David: That’s not a dog. That’s a cat.

In this context, Richard is using the term “dog” to make a claim about what the animal is. Similarly, David is using the terms “dog” and “cat” to make a claim about how things stand with the animal. The truth of Richard and David’s assertions hinges solely on what the animal in the dark room actually is. This is a linguistic disagreement insofar as it is a disagreement over whether or not the term “dog” is true of the animal in the room, and Richard and David can resolve the disagreement simply by getting more information about what the animal actually is.

Now consider a disagreement involving a metalinguistic use of “dog.”

23) *Context: A Cocker Spaniel walks into a room where A and B are sitting.*

Richard: That’s a dog.

David: That’s not a dog. Any dog under 30 pounds is a cat, and cats are pointless.

Assuming that Richard and David have access to identical pools of information about the animal in front of them, this is not a disagreement about whether the animal in front of them belongs to the species-biological class “dog.” Rather, David’s use of “dog” makes clear that he is staking a normative claim about what *ought* to be treated as falling into the extension of “dog,” and where the threshold for doghood ought to be set for the particular conversation.

Similarly, Ludlow (2014) reports a sports radio argument concerning whether racing horse Secretariat was one of the “50 greatest athletes of the 20th century.” Importantly, each interlocutor agreed on all the facts about Secretariat, and the disagreement would not have been resolved by finding some *further* fact about Secretariat (e.g. finding out that he could run even faster) which

would settle things one way or the other. Rather, the disagreement could only be resolved by the interlocutors coming to an agreement on how expansive or restrictive the term “athlete” should be.

Here it becomes clear that metalinguistic negotiations are often driven by background values and normative commitments. For instance, perhaps B believes that we *ought not* to countenance animals under 30 pounds as dogs, since the most appropriate activities for dogs include activities such as hunting, sled-pulling, etc., and sub-30 pound dogs are as useless as cats when it comes to these purposes. Similarly, someone might advocate for a restrictive interpretation of “athlete” to exclude non-human participants like Secretariat on the basis that a more inclusive interpretation might include race cars as “athletes.”

In some cases of metalinguistic negotiation, we can pinpoint an element in the syntax of an expression that is the implicit subject of the negotiation. For instance, Nadia and Kwame might have the following exchange:

24) Nadia: It is *hot* outside.

Kwame: It’s not hot! You ever been to Death Valley? *That’s* hot.

On the analysis of gradable adjectives we are working with, “hot” incorporates a standard parameter whose value is set by context. This means that we have a fairly clear idea of what semantic value in particular is being negotiated here. In effect, Kwame is advocating for the value of the standard parameter for “hot” to be raised to a scalar value which is exceeded by the heat level of Death Valley but not exceeded by the heat level of the current time and location.

Yet it is not always so easy to see what elements of semantic values are being negotiated on the metalinguistic level. “Dog,” “athlete,” and count nouns quite generally do not seem to have syntactic argument places which can be the subjects of metalinguistic negotiations.⁶¹ Yet even if count nouns lack syntactic elements which determine the value of a “threshold” associated with these terms, competent speakers *treat* these terms as if they have such thresholds. In particular, many cases of metalinguistic negotiation involving a count nouns N involve disagreements about whether or not the proposed referential content of N has *enough* of the N-specific properties that all Ns ought to have. So in exchange 23), David stipulates that how much the animal in the room weighs is relevant to the question of whether it is a dog or not. Similarly, Ludlow’s example concerning whether Secretariat is an “athlete” or not seems to be over whether the property *human* is relevant or decisive for ruling individuals into or out of the term’s extension. Call these thresholds which speakers honor in practice *property thresholds*.

One way of incorporating the idea that terms are associated with property thresholds would simply be to define count nouns as denoting a three-place relation among individuals (*i*), a range of properties (*R*), and a threshold (*T_p*) specific to the property *P* being attributed:

$$P = \langle i, R, T_p \rangle$$

In other words, a given individual *i* has to possess “enough” of the *P*-relevant properties in *R* in order to have property *P*, and what determines whether *i* indeed has enough of these *P*-relevant

⁶¹ As we will see later, I think that many count nouns actually *do* have these manipulable elements in their syntax. Here we proceed on the assumption that they do not, in order to show that the thesis of meaning underdetermination does not depend on the thesis of semantic context-sensitivity.

properties is whether or not i is judged to meet the threshold T corresponding to P . If the individual does not meet this threshold, predicating “ P ” of the individual returns a false sentence, and if it does meet this threshold, predicating “ P ” of the individual returns a true sentence. Here we can see that there are at least two dimensions of linguistic meaning up for negotiation. First, there could be negotiations over the properties that ought to be included in R , and which of those properties are most relevant or necessary for a given predication. Second, there could be negotiations over what counts as having “enough” of the properties in R .

A number of considerations speak in favor of this property threshold analysis of count nouns. First, it is a plausible empirical hypothesis in its own right. The idea that lexical items could be associated with lexeme-specific property thresholds is not at all far-fetched. Indeed, as we have seen, it is a working assumption within contextualist approaches to the meaning of gradable adjectives. The conventional meaning of a gradable like “tall” associates it with a scale for *height* rather than a scale for, say, *temperature*. To this extent, from the semantic theorist’s perspective, we get at least part of the meaning of gradable adjectives for free.⁶² Of course, if we endorse meaning underdetermination, the “conventional meaning” of a count noun will do much less work in determining linguistic meaning than more mainstream theories assume. Yet assuming that a suitable explanatory substitute can be found-- perhaps couched in terms of psychological facts about what most speakers of L *take* expressions e_1, \dots, e_n to mean-- positing a link between expression lexicalizations and property thresholds is not at all controversial.

Second, the analysis intuitively captures what is going on in at least some metalinguistic negotiations, especially in contexts where such negotiations don’t seem to target a particular

⁶² Glanzberg (2014) convincingly argues that semantic theorists *must* help themselves to such basic meaning assumptions in getting their theories off the ground.

contextual parameter. Yet perhaps most importantly, it seems to capture what is going on when interlocutors negotiate over the referential content of *racialized* vocabulary. Consider again Jennifer Saul's negotiation over the extension of "immigrant" (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)

As we saw in Chapter 3, it is not clear just *which* subset of foreign-born people who live in the UK "immigrant" picks out in this context.⁶³ Yet what is clear is that Saul seems to lack *whatever* property would be sufficient for ruling her into the extension of "immigrant" for the purposes of these linguistic exchanges.

Finally, note that pairing the thesis of meaning underdetermination with this hypothesis concerning property thresholds gives us a more informative characterization of racialized terms than we have considered so far. We can now define racialized terms as *predicates which include race-specific properties as central elements of their property thresholds at a context*. When race-specific properties are seen as especially relevant to the determination of the property threshold for a term "X", individuals bearing those properties are seen as especially appropriate candidates for belonging in the extension of "X."

We have now given independent conceptual and empirical motivation to our core hypothesis: that the linguistic meaning of racialized discourse is subject to race-specific shifts in

⁶³ Saul has given a more thorough account of the "shifty" nature of "immigrant" in her (2018.)

meaning-standards. We have identified two possible mechanisms for such shifts. First, some racialized terms contain syntactically represented “contextual parameters” whose values shift with features of the context. Second, where racialized terms lack such a parameter, they can either have i) malleable referential contents which are determined by racial features of the context (Chapter 3), or ii) are associated with negotiable “property thresholds” given by the standing meaning of the term.

4.3 Social Perspectives Revisited

So far, we’ve seen that the standing meaning of a lexical item can be “shifty” in two ways. First, standing meaning can be context-sensitive. Second, standing meaning can be underdetermined. Each of these features of linguistic meaning allows the referential content of the lexical item to shift from context to context. Identifying a lexical item as context-sensitive requires us to look for elements of its logical form that may be sensitive to context. Identifying a lexical item as underdetermined requires us to attend to social contestations and negotiations over its meaning. Since standing meaning has such potential for plasticity, we need an account of how interlocutors manage to *converge* on “similar” interpretations of lexical items. That is, speakers and hearers must have a general grasp on the contextually relevant referential content of an expression in order for communication to go smoothly. What explains this convergence?

I have argued that with respect to at least some types of vocabulary, *social perspectives* are what explain our ability to converge on an interpretation of these referential contents. Recall Elisabeth Camp’s (2013) helpful characterization of a perspective:

Above all, perspectives are ongoing dispositions to structure one’s thoughts, along at least two dimensions. First, a perspective involves dispositions to notice and remember certain types of features rather than others, so that those features are more *prominent* or

salient in one's intuitive thinking, and have more influence in determining one's classifications. Second, a perspective involves dispositions to treat some classes of features as more *central* than others, in the sense of taking those features to cause, motivate, or otherwise explain many others. (Camp 2013: 335-336)

This discussion of context-sensitivity and meaning underdetermination gives us two possible roles for social perspectives. Social perspectives can play a role in determining the meaning of E by i) resolving the value of context-sensitive arguments represented in the syntax of E and ii) by lending “determinacy” to the meaning of E in the context of linguistic exchanges. In the next few sections, I will show that some racialized terms have non-obvious context-sensitive elements in their logical form, and that social perspectives are what fix the values of these elements. This section, however, focuses squarely on making sense of the second sort of role for social perspectives.

We have defined racialized terms as terms which incorporate racialized properties as “central” components of their property thresholds. Such properties are more or less central or peripheral components of the property threshold depending on the nature of the social perspective “in play” during the linguistic exchange. In short social perspectives play a role in determining what *property threshold* seems most relevant or appropriate at a context. This hypothesis is supported by empirical findings in psycholinguistics which suggest that the ability to navigate between interpretive perspectives is an intimate part of linguistic competence. Moreover, it seems that humans are particularly good at navigating among interpretive perspectives which are tied to certain social roles. For instance, Barsalou and Sewell (1984) and Barsalou (1987) show that subjects' judgments about what counts as a *typical* or *atypical* member of a category systematically shifts depending on the nature of the perspective being taken up (their term is “point of view.”) In their quaint formulation of the hypothesis:

Since categories often enter into comprehension, retention, and social decision making, their representation may well vary across points of view. For example, if a hippie uttered, “I cooked myself a great dinner last night,” a listener might construct a representation of things to cook for dinner from the hippie’s point of view in which vegetarian dishes were typical and meat dishes were atypical. If the speaker were a redneck, however, the listener might construct a different representation of the category in which meat dishes were typical and vegetarian dishes were atypical. (Barsalou and Sewell 1984: 5)

The subsequent studies find evidence that subjects’ typicality and atypicality membership judgments for a variety of taxonomic (e.g. *birds, crimes, occupations, tools, weapons*) and goal-derived (e.g. *important events in history, important goals in life, kinds of crises, things people worry about*) categories significantly and systematically overlap when those subjects are asked to make such judgments from a variety of cultural (*African, Chinese, American, French*), social role (*hippie, redneck, housewife, businessman*) and abstract (*the average American citizen*) perspectives.

On the basis of these findings, Barsalou and Sewell suggest that interpreters rely on perspectival reasoning in order to discount category-specific information which would be irrelevant or peripheral from the point of view of the perspective being adopted, and to focus on information more relevant or central from that point of view. More specifically, perspective-taking might also aid an interpreter who, with the aid of her background knowledge, is trying to figure out what particular *meaning* of an underdetermined lexical item is in play at a context.

As an example, consider processing the category of food from the point of view of an anorexic. To do this, a person may activate a stereotype for anorexics, which may state that anorexics are compulsive dieters with the goal of being as thin as possible. Although the knowledge base for food may contain a tremendous amount of knowledge, the anorexic stereotype may selectively activate properties relevant to dieting, such as how many calories a food has, how filling it is, and so on... as exemplars become increasingly low in calories and increasingly less filling, they become increasingly typical. (Barsalou and Sewell 1984: 8)

These findings are highly relevant to our claim that the meaning of racialized terms can be determined by social perspectives. For one thing, it suggests that interpreters can come to know that race-specific property thresholds are under discussion simply by recognizing what sorts of social perspectives are in play. An interpreter's understanding of social categories like *thug*, *terrorist*, and *criminal* is mediated by a grasp on a vast store of social knowledge. The question of which individuals are seen as *typical* or *atypical* members of the category depends on the sorts of social perspectives from which those categories are being assessed. In a manner analogous to Barsalou and Sewell's experiments, perhaps interpreters assess these categories from the point of view of *the typical American*, *the White working class*, and so forth. As we saw in Chapter 3, candid racial communication heavily incentivizes its audience to share a perspective from which there are salient, non-trivial connections between racialized properties and membership in these social categories. In some cases, candid racial communicators encourage interpreters to incorporate their category-relevant *racial* knowledge into their representation of who the "criminals," "thugs," and "terrorists" under discussion in fact are. The property thresholds corresponding to terms for these terms become centrally organized around racialized features.

I here do not offer a general account of *how* interpreters manage to defer to or "take on" certain social perspectives in the process of linguistic interpretation. I suspect that we could appeal to different sorts of explanations depending on i) the nature of the perspectival vocabulary involved and ii) the role of that vocabulary in social practices. For instance, in the baseball example discussed in Chapter 2, explaining how players and fans manage to "take on" the umpire's perspective during "strike" calls is rather straightforward. The players i) believe that the umpire is speaking from his own perspective, ii) have beliefs about the rules of baseball that entail that the umpire's perspective is

authoritative over strikehood. Yet we might need a more complex explanation of how perspectives emerge as dominant in the case of racialized vocabulary. (In Chapter 7, I return to this issue via a historical study of how dominant perspectives on racialized criminality emerged.)

Yet even if we lack a general account of how certain social perspectives become dominant, all that matters for our current purposes is that social perspectives can be relevant to the question of what *semantic* interpretation linguistic items receive in context. The empirical evidence reviewed above strongly suggests that this is indeed the case. Now consider a more *a priori* sort of argument for this conclusion.

As we noted in Chapter 1, sometimes the very use of a term indicates the general social perspective being presupposed. The fact that certain “lexical choices” can indicate how the chooser thinks about the phenomenon being described may seem like a truism. Yet as we’ve seen, it’s more complicated to explain why this “truism” obtains. On my account, the referential content of perspectival terms can systematically shift along with the social perspectives which wield or apply those terms. Recall Kwame Ture (born Stokely Carmichael) and Charles Hamilton’s series of racialized examples:

In the wars between the white settlers and the “Indians,” a battle won by the Cavalry was described as a “victory.” The “Indians” triumphs, however, were “massacres.” (The American colonists were not unaware of the need to define their acts in their own terms. They labeled their fight against England a “revolution;” the English attempted to demean it by calling it “insubordination” or “riotous.”) The historical period following Reconstruction in the South after the Civil War has been called by many historians the period of Redemption, implying that the bigoted Southern slave societies were “redeemed” from the hands of “reckless and irresponsible” black rulers... Thus black people came to be depicted as “lazy,” “apathetic,” “dumb,” “shiftless,” “good-timers.” Just as red men had to be recorded as “savages” to justify the white man’s theft of their land, so black men had to be vilified in order to justify their continued oppression. (Hamilton and Ture 2011: 35)

Ture and Hamilton's passage provides some nice example of how dominant social perspectives might play a role in determining the intuitive *truth-values* of sentences containing perspectival vocabulary.⁶⁴ A colonial-era Brit, embedded in colonial-era British ways of seeing global politics, might judge the sentence "There is a revolution going on in America" to be false, since from his geopolitical perspective, organized around the idea that Britain is the source of legitimate colonial authority, the colonists' actions are not a *revolution* against an illegitimate authority, but a *riot* against a legitimate one. Moreover, since the extension of terms such as "revolution" and "riot" are populated by just those things that are *treated* as revolutions and riots by some dominant social perspective, social perspectives hold a great deal of sway over what "revolution" and "riot" can *refer* to at a context. Similar lessons of course apply to what counts as a "massacre" or a "victory," and whether or not the South was truly "redeemed" from Black governance.

Moreover, terms like "lazy," "apathetic," "dumb," "shiftless," and "savages" indicate the influence of a social perspective which selectively foregrounds the negative properties of racialized groups. As I argued in Chapters 2 and 3, the fact that this perspective indeed foregrounds these properties has effects on the truth-conditions of the sentences in which they are embedded. As we saw, this is particularly true of generic sentences: the threshold for *shiftlessness* that Blacks have to meet in order for "Blacks are shiftless" to be true (to an interpreter) differs from the threshold that Whites have to meet.

⁶⁴ This is in fact a rather intuitive view. It is, for instance, commonly taken to be one of the intuitions a semantics for predicates of personal taste must account for (Laserson 2005, Glanzberg 2007, McFarlane 2014). The truth-value of "Goulash is tasty" naturally switches depending on which perspectives we consult: the sentence is false from the perspective of those who do not find goulash tasty, and true from the perspective of those who find it tasty. The truth of the sentence is entirely perspective-dependent, and this perspective-dependence is perhaps traceable to the presence of hidden contextual parameters in "tasty" and other predicates of personal taste (e.g. Glanzberg 2007). See Appendix 1 for more details concerning the semantics of predicates of personal taste.

4.4 Combining the metasemantic hypotheses

Now we can give a first-pass analysis of the interaction between social perspectives, meaning-underdetermination, and context-sensitivity. Since the meanings of racialized terms are often context-sensitive, underdetermined, or both, it is the job of social perspectives to lend referential determinacy to racialized terms. The remainder of this section will explore how social perspectives interact with the property-threshold analysis given in Section 4.2. Sections 5 and 6 will more squarely focus on how social perspectives can interact with contextual parameters represented in the syntax of racialized terms.

Consider the example “terrorist,” and in particular the sentence “Dylann Roof is a terrorist.”

⁶⁵ Assuming that “terrorist” is paired with a contextual property-specific threshold, we represent the sentence as follows:

Sentence: Dylan Roof is a terrorist.

Lexical interpretation: **[[terrorist]]** = $\langle \text{Roof}, R, T_{\text{terrorist}} \rangle$

Of course, there are a variety of properties that could seem particularly relevant to the threshold. But as we will see in Chapter 5, it is plausible that these thresholds contain elements like the following:

Available property threshold elements: MUSLIM, NON-WHITE, ARAB, WHITE, etc.

⁶⁵ Roof is a White Supremacist and mass murderer who on June 17, 2015 shot and killed nine African-American worshippers in a church in Charleston, SC. Despite his clear ideological motivations and reliance on brutal, targeted violence, Roof was never characterized as a “domestic terrorist” by any authoritative body, nor was he so described in the majority of media coverage. Plausibly, this is because Whites are not generally thought of as prototypical examples of terrorists.

On the conception of “semantics” I have defended above, this is all “seen” by the lexical semantics of “terrorist.” The fact that there are multiple possible elements of the property threshold reflects the fact that “terrorist” is semantically underdetermined.

Now for the metasemantic level. The role of social perspectives is to select certain elements of the property threshold as particularly *salient*. As we have seen, a number of things can affect just what perspective is in play, and what sorts of salience relations that perspective “sees.” Chapter 3 made the case that the norms of one’s *discourse community* plays a major role here, Chapter 5 argues that one’s psychology might constrain these perspectives, and Chapter 7 will make the case that perspectives on certain racialized categories are broadly and systematically distributed enough to become “default” perspectives on these categories. Let’s assume here that an interpreter accesses a perspective on “terrorist” that foregrounds the property of *being Muslim*.

Metasemantic Perspective: MUSLIM central (other properties more peripheral)

Lexical Interpretation: **[[terrorist_{MUSLIM}]]**

The Lexical Interpretation of “terrorist” here indicates that the referential content of “terrorist” is constrained by the property *Muslim*. It is as if *Muslim* provides the interpreter with a heuristic which limits the space of “good candidates” for the referential content of “terrorist.” The original sentence “Dylann Roof is a terrorist” is thus determined to be *false* by the interpreter. In the next section, we’ll investigate whether these “constraints” can be located in the syntactic representation of racialized terms.

5. A First-Order Semantics for Some Racialized Terms

The point of dwelling on metasemantics at such length has been to show that the semantic

plasticity or “shiftiness” characteristic of racialized terms can be traced to quite general features of linguistic meaning. Racialized vocabulary is special in the sense that its semantic plasticity is tied to race-specific shifts in meaning-standards, but it is ordinary in the sense that it is subject to the same sorts of meaning underdetermination, context-sensitivity, and perspective-sensitivity that characterize large swaths of non-racialized language as well.

This section shows that the theses of context-sensitivity, meaning underdetermination, and social perspectivalism can be profitably joined with a first-order semantics for some racialized terms. Focusing on examples like “thug” and “terrorist,” I argue that many racialized terms are *inherently generic*. If successful, this argument yields a number of theoretical benefits. First, it allows us to track elements in the syntax of racialized terms whose semantic values can be fixed (directly) by context and (indirectly) by social perspectives. Consequently, it permits us to give an informed diagnosis of why the referential content of racialized terms shifts in the face of race. Second, it allows us to show how interpreters’ understanding of the truth-conditional content of racialized terms can vary depending on whether a racialized property threshold is being observed or not. Finally, treating some racialized terms as inherently generic allows us to generalize about the sorts of psycho-pragmatic effects that *both* Type C generics and racialized terms have on conversational context. We take up that issue in Chapter 6.

5.1 Genericity again

We discussed generics at length in Chapter 2, but it will be helpful to remind ourselves of some of the properties of generic language. Generic language allows human beings to move beyond statements about particular situations, facts, and individuals and to make claims about lawlike

regularities, general states of affairs, and kinds. There are two paradigmatic types of genericity. The first is attributable to the presence of *kind-referring* expressions which pick out kinds of objects rather than concrete individual objects. The second is that of *characterizing sentences* which are used to express general regularities or laws. These two types of genericity frequently overlap. In the following examples, kind-referring NPs are underlined:

KIND-REFERRING GENERICITY

- 25) The potato was first cultivated in South America.
26) Long ago, my ancestors hunted the buffalo.

CHARACTERIZING GENERICITY

- 27) John [usually] smokes after dinner.
28) The anteater is between 14 inches and 7 feet long.

OVERLAP:

- 29) Dogs bark.
30) Ducks lay eggs.

25) and 26) are statements of fact about potatoes and buffalo qua generic, *open-ended classes* of objects. 27) and 28), on the other hand, characterize what concrete particulars (John and the arbitrary anteater) are like across an indeterminate range of situations, and as such are generic statements about the regularity with which John smokes and the range of possible sizes for the arbitrary

anteater. Finally, 29) and 30) seem to predicate general dispositional properties (that is, properties that are or can be displayed across a range of situations) to dog-kind and duck-kind.⁶⁶

In addition to these two types of genericity, there is evidence for a third type of genericity due to the presence of “individual-level predicates.” Linguists (Carlson 1977, Carlson and Pelletier 1995) have long insisted on a distinction between *individual-level* and *stage-level* predicates.

Individual-level predicates pick out fairly stable or permanent features of objects, while stage-level predicates pick out more fleeting or temporary properties. For instance, the fact that I am walking in the park right now is a stage-level fact about me, since I have that property over a relatively short amount of time. On the other hand, the fact that I am over six feet tall is an individual-level fact about me. In the absence of serious accidents capable of affecting my height, I will continue to have the property of being over six feet tall.

Consequently, within the class of i-level predicates are predicative NPs and adjectives that denote stable or enduring properties of individuals, and within the class of s-level predicates are VPs and adjectives which denote fleeting or non-permanent properties. Consider an intuitive list:

Stage-level predicates	Individual-Level Predicates
...is speaking Italian ...is drunk ...is smoking ...speaks Chinese	...knows Italian ...is an alcoholic ...is a smoker ...is Chinese

I-level and S-level predicates

Figure 4

⁶⁶ This is somewhat contentious, since many theorists hold that in characterizing sentences, the GEN operator quantifies over individuals, not kinds. However, the contrary hypothesis also has a purchase in the literature (see Teichman 2017). In any case, not much hinges here on whether the characterizing sentences I am interested in indeed include kind-referring genericity.

Like any intuitive distinction, this admits of borderline cases. For instance, “Black-haired” seems to be a candidate for an individual-level predicate, since hair colors are fairly constant. We might be tempted to treat it as a stage-level predicate, however, if it became fashionable to dye one’s hair a different color every day. Consequently, i-level and s-level predicates are also distinguished by linguistic tests.

For instance, only (some) S-level adjectives can be affixed to the coda position in there-constructions, but I-level adjectives cannot be so affixed.

31) *S-level*: There are officers available/standing/talking/drunk/ready...

32) *I-level*: ?? There are officers Black/intelligent/foolish/brave/knowning/thuggish...

Another piece of evidence is that such noun phrases cannot be embedded in the small clauses of “perception sentences” :

33) *S-level*: I saw(/heard) Tyrone eating/drunk/speak Chinese

34) *I-level*: ?? I saw(/heard) Tyrone tall/a doctor/Chinese.

Finally, consider that i-level predicates generally cannot be modified by locative and temporal phrases such as “in,” “on,” “before,” “after,” and the like, while s-level predicates can be so modified:

35) *S-level*: Nadia eats sushi in her car/on the lawn/before noon.

36) *I-level*: Nadia is a doctor in her car/on the lawn/before noon.

S-level and I-level predicates thus behave differently in fairly systematic ways. What semantic difference might explain this differential behavior?

One proposal (Kratzer 1995) is that s-level predicates contain a Davidsonian argument *l* picking out space-time locations, and that i-level predicates lack such an argument. This sits well with the intuition that as descriptions of particular events, s-level predicates pick out objects and properties which are specifically spatiotemporally located, while i-level predicates seem to pick out properties which are not located “at” any particular spatiotemporal context.

S-level predicates	I-level predicates
“Manon <u>is dancing</u> on the lawn.”	“Manon <u>is a dancer</u> .”
[dancing (Manon, l) & on-the-lawn (l)]	dancer(Manon)

Kratzer’s semantics for S-level and I-level predicates

Figure 5

And indeed, this could explain why the interpretation of s-level predicates is sensitive to shifts in the spatial and temporal context, while that of i-level predicates is not: while I may be *drunk* now, I will be sober in a few hours, and while I am *available* to meet with students on Tuesday when I am in California, I am not so available now in Michigan. On the other hand, regardless of where or when I am, the properties of *being six feet tall* or *being bald* “follow me around,” at least for a rather substantial portion of my lifetime, in a way that is largely insensitive to my space-time location. In other words, as Gennaro Chierchia puts it, i-level predicates “ascribe tendentially stable properties to their arguments. It seems that one can say of an argument with an i-level property P, ‘once a P, always a P.’” (Chierchia 1995: 198)

Yet Kratzer’s proposal effectively treats a difference in the surface behavior of two sorts of predicates as evidence of a deep syntactic difference in the argument structure of i-level and s-level predicates. Perhaps we can account for this difference in surface behavior in a more theoretically parsimonious manner. Correspondingly, Chierchia (1995) himself has proposed a different account of the semantic differences between I-level and S-level predicates. Rather than claiming that I-level predicates *lack* an argument which S-level predicates have, he argues that *both* species of predicate contain a Davidsonian argument denoting *situations*. What distinguishes I-level and S-level predicates has to do with the characteristic range of situations over which each type of predicate tends to apply. In particular, the arguments of I-level predicates are bound by a *generic* operator, rather s-level predicates are bound by more familiar quantificational operators. The difference between i-level and s-level predicates, then, is not treated as evidence of two different syntactic argument structures, but as evidence of a difference in the generality of the quantifier affixed to such structures.

As is standard, Chierchia treats GEN as a two-place operator which relates two semantic constituents to one another: i) bound variables existing in a range of situations in the restrictor and ii) i-level properties in the matrix. For instance, an i-level predicate like “altruistic” receives the following semantic gloss:

$$[[\mathbf{altruistic}]] = \lambda x \text{ GEN}_s [\text{in}(x, s)][\text{altruistic}(x)]$$

This analysis makes clear that the truth of a sentence such as “x is altruistic” depends on x acting in an altruistic manner, or at least having a stable disposition to act altruistically, across a general range

of circumstances.⁶⁷ On the other hand, the truth of a sentence like “is running” only depends on x’s running in *at least one* situation (namely, the one being described):

$$[[\mathbf{running}]] = \lambda x \exists_s [\text{in}(x, s)][\text{running}(x)]$$

At the level of lexical semantics, the core difference between i-level predicates and s-level predicates is that the latter incorporates an existential quantifier, while the former incorporates a GEN operator.

S-level predicates	I-level predicates
“Manon is dancing on the lawn.” $\exists_s [[\text{in}(\text{Manon}, s)][\text{dancing}(\text{Manon})][\text{on-the-lawn}(\text{Manon})]]$	“Manon is a dancer” $\text{GEN}_s [\text{in}(\text{Manon}, s)][\text{dancer}(\text{Manon})]$

Chierchia’s semantics for S-level and I-level predicates

Figure 6

One thing that lends independent motivation to Chierchia’s account vis a vis Kratzer’s is that the former account nicely explains several interesting commonalities between the behavior of generics and i-level predicates. First, it helps make sense of the idea that an individual can possess an i-level property even if it does not instantiate that property in *all* or even *most* situations. Recall that the truth of generic sentences, unlike the truth of the kind-wide quantifications that are their cousins, is compatible with the existence of counterexamples to those sentences: “birds fly” is not falsified by

⁶⁷ Of course, determining the extent of this “range of circumstances” isn’t something specified by Chierchia’s semantics. Given our discussion of the notoriously slippery truth-conditions of generics in Chapter 2, this is not surprising. Yet there might be a natural interface between Chierchia’s proposal and the account of property thresholds presented earlier in this chapter. Whereas Chierchia’s lexicalizations of s-level predicates elucidate the basic truth-conditions of these predicates, my account of situation thresholds demonstrates the particular way in which these basic truth-conditions are context-sensitive.

the fact that penguins do not fly, but “all birds fly” is falsified. Similarly, someone’s being an *alcoholic* or a *murderer* in a situation is compatible with their not actually drinking/seeking alcohol or murdering anyone in that situation. Contrast this with the example of an s-level predicate like “standing up.” An individual’s *standing up* in a situation is, trivially, incompatible with their *not standing up* in that situation.

Second, both i-level predicates and generics tend to be understood as describing states of affairs and properties that hold more or less *universally*, independently of particular *locational* contexts. Generics, for example, unlike quantified statements, do not *domain-restrict*; that is, the GEN operator does not automatically incorporate information from the context which can be used by an interpreter to disambiguate the sentence. For instance, if I go to a party and say, “everyone is here!,” I clearly do not mean (and am not interpreted as meaning) that everyone *in the universe* is here, or that everyone *in the city* is here. Rather, “everyone” means some contextually relevant subset of all possible partygoers (e.g. my friends and family), but that isn’t something that I have to explicitly state. The quantifier automatically restricts our attention to that relevant subset. On the other hand, generic language does not show these signs of contextual domain restriction. For instance, imagine that I go to a reservation for injured animals who cannot survive in the wild. I notice that all the squirrels on this reservation only have three legs. In this case, 37), but not 38), would be a true and felicitous statement:

37) Wow, every squirrel has three legs!

38) ?? Wow, squirrels have three legs!

38) sounds strange precisely because the generic seems to say something about what squirrels are like in general outside of special contexts, whereas 37) is perfectly fine, since it implicitly restricts the domain to squirrels within a special context.

GEN thus encourages interpreters, in the absence of auxiliary information, to interpret generic expressions as taking broadest possible scope, independently of location. Consequently, the presence of a GEN operator in the semantics of i-level predicates, may explain why predicative noun phrases such as “is an alcoholic” or “is a doctor,” for instance cannot be modified by locative phrases, while s-level predicates can be so modified (Chierchia 1995).

39) S-level habituals: Tyrone eats/smokes/listens to music in his car.

40) S-level adjectives: Tyrone is happy/sad/sleepy in his car.

41) I-level noun phrases: ?? Tyrone is an alcoholic/a doctor in his car.

Third, the presence of a GEN operator in the semantics of I-level predicates may explain why the properties described by i-level predicates seem to interpreters to be *deep or lasting*, whereas those properties described by s-level predicates seem to be *superficial or fleeting*. In short, in addition to providing a semantics for i-level predicates, it may help explain why human interpreters *take* i-level predicates to describe stable, location-independent properties in the first place. Generic language, as we have seen, is intimately linked to our tendency to *essentialize* certain kinds, properties, and features of the world. The presence of a GEN operator in the semantics of many i-level predicates thus might take us some distance in understanding *why* i-level properties are the sorts of things that are

taken to be more central or essential to the identity of an individual, whereas s-level properties are taken to be more accidental or peripheral.⁶⁸

Consider one final source of support for the claim that i-level predicates are inherent generics. Recent evidence suggests that both children are more likely to treat a property as more stable and context-independent when it is lexicalized in the form of a *label*. Gelman and Heyman (1999) found that labelling an individual increased children's assessments of whether the person would display the property across a wide variety of environments. For instance, labelling Rose as a "carrot-eater" rather than describing her as someone who "eats carrots whenever she can" increased children's confidence that Rose had been a carrot-eater in the past, would remain a carrot-eater in the future, would remain a carrot-eater despite the opposition of her family, etc. All of the labels tested were i-level predicates, and all the verb-based descriptions were s-level predicates. This provides a good deal of support for Chierchia's suggestion that i-level predicates are treated as more stable and inductively richer than s-level predicates. Moreover, Waxman (2010) found that children generally regard novel properties (e.g. a love for a game called "zaggit") as distributed across the category of *persons*, with no particular bias to differently gendered or racialized persons. However, describing diversely racialized and gendered individuals *as* belonging to a *novel* labelled category such as "Wayshan" increased children's confidence that Wayshans were fundamentally distinct sorts of people who shared non-obvious, novel properties (e.g. a love for a game called "zaggit.")⁶⁹

⁶⁸ This claim shouldn't be overstated. Social context and common sense conception will constrain the sorts of categories are in fact essentialized, and so the answer to the question of which kinds and properties end up being essentialized will depend on much more than the question of what kinds and properties are described via inherent generics.

⁶⁹ Different race/sex individuals were described via different lexicalizations (e.g. "Wayshan" for Black females, "Dappo" for White females, etc.) The point of course was to see whether the lexicalization of the novel property itself was prompting the belief that a novel individual Wayshan would have a given property. Interestingly, children did not rely on race-membership as a guide to property-instantiation when individuals were *not* labelled as Wayshans, Dappos, etc., indicating that race-knowledge alone was not prompting these beliefs. Consequently, this also supplies support for my hypothesis that racialized discourse can serve as particularly potent "triggers" for representations of racialized categories.

Obviously there is a family resemblance here between the apparent cognitive effects of labelling and those of generic interpretation. The fact that children treat novel labels as so inductively potent, sometimes to the exclusion of their world-knowledge, is strong support for Chierchia’s hypothesis that i-level predicates are generics. We will return to this connection between inductive potentials and generic language when we broach the topic of *essentialism* in Chapters 5 and 6.

6. Racialized Terms as Inherent Generics

Chierchia’s semantics has obvious implications for our current project. Many of the particular terms we are tracking-- “thug,” “terrorist,” “immigrant,” and the like-- are i-level predicates, and *ex hypothesi* inherently generic. On this analysis, the lexical entries of many racialized terms will be of the following form:

$$[[\mathbf{thug}]] = \lambda x \text{ GEN}_s [\text{in } (x, s)] [\text{thug } (x, s)]$$

$$[[\mathbf{terrorist}]] = \lambda x \text{ GEN}_s [\text{in } (x, s)] [\text{terrorist } (x, s)]$$

$$[[\mathbf{immigrant}]] = \lambda x \text{ GEN}_s [\text{in } (x, s)] [\text{immigrant } (x, s)]$$

This analysis can potentially shed some more light on the mechanisms that drive race-related shifts in meaning-standards. As we will see, Chierchia’s proposal allows us to locate semantic constituents within racialized i-level predicates which can be manipulated in context and targeted by semantic negotiations.

To see this, let’s work through some general implications of Chierchia’s semantics. Chierchia’s semantics entails that the truth-conditional content of an i-level predicate is determined by the semantic contribution of the GEN operator. More specifically, for any given application of an

i-level predicate “X” to an individual, there must be some answer to the question of how *generally* an individual must instantiate the ascribed property in order to satisfy “X” at the context. Call the answer to this question a *situation threshold*. Consider an example:

42) Bob is an alcoholic.

42) is true iff Bob is an alcoholic. Whether or not 42) is true hinges on whether Bob satisfies the term “alcoholic.” On Chierchia’s semantics, whether Bob satisfies the term “alcoholic” hinges on whether Bob generally instantiates whatever property or disposition “alcoholic” picks out in a sufficiently general range of situations. That is, 42) is true if Bob meets the *situation threshold* for “alcoholic.”

Similar lessons apply to the case of racialized terms like “thug,” “terrorist,” and “immigrant.” Being a thug, for instance, is not a matter of occupying fleeting or non-permanent stages. A “thug” is a person whose thuggish disposition and behavior is instantiated over a general *range* of situations, and the fact that a thug does not behave thuggishly in a particular situation is no evidence for their not having that character trait. What Chierchia’s semantics illustrates is that the truth-conditions of many (i-level) racialized terms depend on there being some general range of situations which distinguishes true from false predications involving racialized vocabulary.

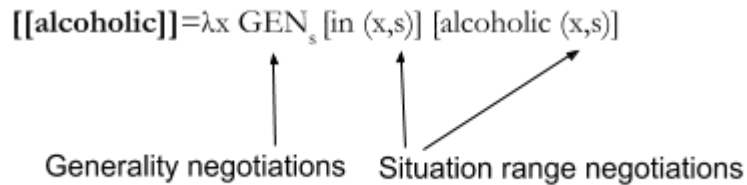
Of course, determining what “general range” of situations involving an individual and a use of “thug” is in fact *sufficient* for ruling an individual into the extension of the term at a context is not always clear. Since situation thresholds are a general feature of Chierchia’s semantics (though not identified as such)⁷⁰, there is often a good deal of indeterminacy concerning whether an individual satisfies an i-level predicate. For some terms, it’s clear what the value of the situation threshold is.

⁷⁰ The notion of a situation threshold is my own gloss on Chierchia’s semantics. I am not certain that Chierchia himself would agree with this interpretation.

For instance, anyone who counts as a “murderer” or a “rapist” must only have instantiated these corresponding properties in one actual situation. Other terms are associated with high thresholds-- those appropriately described as “omniscient” and “omnipotent” must not fail to be all-knowing or all-powerful in *every* relevant situation.

Yet the vast majority of our terms have murky situation thresholds. “Red” things, “annoying” people, and “dangerous” drugs do not seem to have principled minimal or maximal thresholds which strictly govern the application of these terms. Consequently, in using and interpreting these murky terms, speakers often rely on an intuitive background grasp of the sorts of situations which count in favor of the description’s accuracy and which count against it. What is important for our purposes is that when negotiations and contestations over i-level predicates arise, Chierchia’s semantics allows us to see what particular semantic elements are at the heart of this negotiation.

I will here discuss two possible ways in which the situation threshold can be negotiated. The first concerns negotiations over what *types* of situations are properly included in the situation threshold. The second concerns negotiations over how *generally* a property must be displayed range of situations. These negotiations thus target two different elements in the lexical semantics of i-level predicates: the *situation range*, and the GEN operator:



Let’s take an example of a negotiation over the situation range first. Consider a famous example from Sellars (1956). John owns a necktie shop, and “has never looked at an object in other

than standard conditions.” He can competently classify the neckties in his shop as “green,” “red,” “purple,” and so on. Then his shop installs electric lighting, which makes the color of the neckties appear somewhat different. One day, Jim comes in to buy a necktie:

"Here is a handsome green one," says John.

But it *isn't* green," says Jim, and takes John outside.

"Well," says John, "it was green in there, but now it is blue."

"No," says Jim, "you know that neckties don't change their color merely as a result of being taken from place to place."

"But perhaps electricity changes their color and they change back again in daylight?"

"That would be a queer kind of change, wouldn't it?" says Jim.

"I suppose so," says bewildered John. "But we *saw* that it was green *in there*."

No, we didn't see that it was green in there, because it wasn't green, and you can't see what isn't so!"

(Sellars 1956: 143).

Sellars' point is that the notion of “being green” is logically prior to the notion of “looking green.”

This is to the detriment of sense-datum theories, which reverse this relationship of priority. My point is simply that John and Jim seem to be working with different conceptions of what the appropriate situation threshold is for the term “green.”⁷¹ What is implicitly under discussion is what sorts of situations would render the following sentence *false*:

43) This necktie is green.

According to John (at least at first), 43) is true inside the shop, and false outside of it. 43) is false outside the shop because “electricity” has changed the color of the neckties. John thus seems to be honoring a situation threshold for green which generalizes over situations involving *different kinds of light*. Strictly speaking, John's lexicon splits “green” into a number of entries: there is the property “green” picks out in daylight situations, the green of night-time situations, the green of electric light

⁷¹ Note that while “green” is a gradable adjective, John and Jim do not seem to be disagreeing over “how green” the necktie is. Rather, they are disagreeing about whether it is green *at all*. This indicates that this particular negotiation is over the situation threshold, not the standard parameter.

situations, and so on. None of this needs to be explicit, or articulable to John. The difference comes through in his linguistic behavior. On the other hand, Jim believes (43) to be true regardless of the context in which the tie is viewed. He effectively points out that John honors an absurd situation threshold for “green”-- after all, “neckties don’t change their color merely as a result of being taken from place to place.” John thus advocates for a less fragmented interpretation of the situation threshold. “Green” denotes a stable color property that obtains across a general range of “normal” situations, perhaps.

Now consider instances of negotiation over *generality*, and particularly over the GEN operator. In these sorts of negotiations, what is under discussion is whether an individual instantiates a given property in a *sufficient* range of situations. On this view, Blacks and Muslims will intuitively be taken to instantiate whatever properties are assumed to be “thug/terrorist-consistent” in a *broader* range of situations, or perhaps *fewer* situations will be required for these individuals to meet the operative threshold for “thug” and “terrorist.” By the same token, the threshold corresponding to these lexical items will be assigned a more restrictive semantic value when Whites and non-Muslims are under discussion. For some racialized terms, we can locate this restriction within the GEN operator.

Finally, note that the manipulable elements in the syntax of i-level racialized terms-- the GEN operator and the situation range-- are manipulated against the backdrop of a certain interpretation of the property threshold. This shows us that no part of this rich metasemantic story is idle: the notion of a situation range is parasitic on the notion of a property threshold, which is in turn parasitic on the notion of a social perspective.

7. Conclusion: Thresholds and Defeasibility

This chapter has pursued two goals. First, it has proposed an ambitious, global account of how racialized meanings can be expressed via language which does not seem obviously racialized. Second, it has fleshed out the semantic and metasemantic assumptions that have been driving my argument so far, in the interest of showing how theoretical and empirical advances in the philosophy of language and linguistics could profitably be applied to a type of racialized discourse which has hitherto avoided attention from these disciplines.

Of course, critics might insist that we don't need a baroque metasemantics to explain race-specific meaning-shifts when we have the resources of pragmatics, or plausible psychological stories about how individuals acquire and update racialized beliefs via racialized discourse. Yet on the inflationary conception of semantics I have pursued here, resources usually coded as "pragmatic" or "epistemic" under the traditional division of labor also bear on our understanding of the nature of linguistic meaning. I thus see these critics as endorsing projects which are complementary to my own, not in competition with it.

Moreover, two things especially speak in favor of the semantics I have assembled here. First, that account has wide-ranging and exciting applications to a wide variety of issues surrounding the study of how language can be used to distribute racialized representations through language. As we will see in Chapters 5 and 6, the first-order semantics for (some) racialized terms nicely captures some aspects of how people defaultly *think* and *reason* about racialized categories. The second thing that speaks in favor of my account is that it provides *multiple* solutions to what we identified as the *threshold problem* and the *defeasibility problem*. We close by discussing a number of possible solutions.

7.1 *The Threshold Problem*

We have seen that there are several linguistic mechanisms that speakers can exploit in negotiating the semantic boundaries of their terms. To the extent that these mechanisms are also incorporated in the terms which we have identified as racialized, students of racialized discourse can avail themselves of sophisticated semantic tools to track race-specific shifts in meaning-standards.⁷² In explaining why differently racialized individuals and groups seem to face different standards for the application of a term, we may appeal to the fact that racialized terms are *socially perspectival* and *context-sensitive*.

Chapters 2 and 3 argued that *race-specific* features of the linguistic and non-linguistic context can affect both the standing meaning and the referential content of racialized terms. Showing that a given expression or expression type is semantically context-sensitive usually requires us to identify some element in the syntax of that expression whose contribution to truth-conditional content is sensitive to features of the non-linguistic context. This chapter has given content to this idea by pointing to possible elements in the syntax of some racialized terms whose semantic values could be determined by elements of the “racial” context.

We thus have a straightforward answer to the question of why certain racialized individuals and groups face different thresholds for the application of a term. While we could see this as a systematic feature of the *use* of some terms, this chapter has provided the resources for thinking of these differential thresholds as systematic features of the *meaning* of these terms as well.

⁷² Different tools will likely be needed for different lexical types, of course. While this project is most centrally concerned with how racialized discourse can be better understood through an engagement with generics and count nouns, we may need different semantic tools to account for the sorts of meaning-shifts characteristic of racialized *verbs* (“to riot,” “to immigrate,” “to infest,” etc.) There is a sense in which my goals are programmatic: I’ve shown how fine-grained linguistic analysis and racialized discourse could be relevant to one another in a set of specific cases, and I would be happy to see philosophers and linguists continue with this sort of work.

7.2 *The Defeasibility Problem*

As we saw in Chapter 3, racialized terms like like “thug” seem to carry the information that a racialized individual or kind is under discussion in *some* contexts, while they do not seem to carry such information in other contexts. As we saw in Section 1 of this chapter, this is not a feature shared by many other race-specific terms. Those who favor pragmatic approaches often take this cancelability to suggest that if race-specific meanings were somehow lexically encoded within racialized terms, sentences which explicitly “cancel” those contents should sound contradictory:

44) Dylan Roof is a terrorist, but he is not a Muslim.

45) Jens is a non-Latino immigrant from Sweden.

The fact that these sentence are intuitively fine is taken to be evidence that pragmatic approaches fare better.

Yet if the semantic and metasemantic treatment of racialized terms I have offered in this chapter is on the right track, there are many ways to model this defeasibility without appealing to pragmatic effects. The canceling clauses make clear that a term is being subjected to a race-specific shift in meaning-standards. Such shifts could be attributable to i) the lexical structure of inherently generic i-level predicates, or ii) the fact that the meanings of these terms are underdetermined. In closing, we’ll consider each hypothesis in turn.

7.2.1 *Defeasibility due to shifts in generality*

A well-known feature of generic language is that it is closely related to *non-monotonic* patterns of inference. In general, if one knows that *Ks are F*, one is not allowed to conclude that any *particular*

k within the set of Ks is F . In fact, given *any* set of good inferences that can be drawn from a generic sentence of the form Ks are F , adding more information about kind K can make previously valid inferences invalid.⁷³ Suppose that every *bird* we've encountered so far can *fly*. We might form the belief that if x is a bird, then x flies. Yet then come across kiwis, ostriches, penguins, and are told that even though these animals don't fly, they are still birds. We might then qualify our original belief: if x is a bird, x flies *unless* it is a penguin, kiwi, ostrich, etc. What at first seemed to be a valid inference was seen to be invalid once more premises were added.

In a parallel fashion, if racialized terms are inherent generics, we should expect interpreters to grasp that race-specific individuals are under discussion only if they are not in possession of background knowledge which would render that inference invalid. Yet interpreters are in possession of this information much of the time. In fact, as we saw in Chapter 3, speakers frequently need to put a good deal of effort into manipulating the context so that the racialized meaning comes across. Consequently, we should think of defeasibility not as the exception, but more as a rule for racialized terms, especially those which have inherently generic semantic structure.

7.2.2 *Defeasibility due to meaning-underdetermination*

If lexical meanings are truly underdetermined, no lexical item has a privileged sense independently of the way it is used in concrete linguistic exchanges. This has a few upshots. First, it means that the “defeasibility test” is no longer a reliable discriminator of when a meaning paired with a lexical item L is “semantic”/linguistically encoded or “pragmatic”/post-semantic. If meaning

⁷³ This non-monotonicity is not characteristic of a special class of generics which Khemlani, Leslie, and Glucksberg (2009) call “quasi-definitional generics”: “poodles are dogs,” “bachelors are unmarried men,” “gold things are metal,” etc..

is underdetermined, we should expect *all* lexical items to be connected with their standing meanings only in a defeasible way. For instance consider the following sentence:

46) I have a husky, but she's not a dog.

On a natural reading, the second clause seems to explicitly contradict what is said by the first. Since it seems false that one could *both* have a husky *and* have that husky not be a dog, the sentence must be a contradiction. Since the second clause does not successfully cancel the first, this suggests that *dog* is “part of the content” of “husky.” Yet if meaning is underdetermined, the lexical items in the sentence could have any number of meanings depending on the character of the linguistic exchange in which the sentence is uttered. So, for instance, where it is agreed that “dog” is shorthand for *ugly nasty old cur*, the sentence is perfectly felicitous.

So, if meaning is underdetermined, the defeasibility test does not show very much about how much of a term's meaning is linguistically encoded or not. The obvious upshot here is that the acceptability of sentences such as “Dylann Roof is a terrorist, but he is not Muslim” do not automatically show that racialized terms have no racial “content.” Whether they do or not, I take it, rests on facts about the linguistic exchange in which these sorts of sentences are used.

Chapter 5. Essentializing and Moralizing Racialized Conduct

Abstract

This chapter shifts from an examination of the semantic properties of racialized *language* to an examination of the contents of racialized *thought*. Whereas Chapters 2-4 were concerned with how interpreters process certain race-specific linguistic strings, this chapter is concerned with how interpreters *mentally represent* the relationships between racialized groups and their perceived category-typical behaviors.

Within the broader context of the dissertation, this chapter marshalls evidence that race-specific shifts in conduct-standards are widespread, and it provides a hypothesis about the psycho-social mechanisms that underlie these shifts. The hypothesis consists of three claims. First, the ambiguously racialized categories we have been tracking-- criminals, immigrants, terrorists, welfare recipients, and the like-- are particularly susceptible not only to being racialized, but to being *essentialized*. Second, the notion of “essence” that is taken to *explain* the conduct of category members shifts depending on how they are racialized. Finally, group members are taken to be more or less *responsible*, and therefore more or less susceptible to praise and blame, depending on how they are racialized. There have been a handful of systematic attempts to unite psychological data concerning essentialism, racial bias, and both causal and moral forms of attribution under one roof. I discuss two of these systematic attempts, Pettigrew’s (1979) “Ultimate Attribution Error” (UAE) and Leslie’s (2017) “Supreme Attribution Error” (SAE). I show that both are confronted by problems sufficient to call for a new hypothesis. In particular, neither account does sufficient work to show that widespread essentialism concerning a racial category is consistent with widespread responsibility attributions for having traits and displaying behaviors that are perceived as rooted in a group

essence. The final section of the chapter thus proposes, as a working hypothesis, a new sort of attribution error which does precisely this: The Moralistic Attribution Error.

To consider every member of a group as endowed with the same traits saves us from the pains of dealing with them as individuals... a belief in essence develops. There is an inherent Jewishness in every Jew. The "soul of the Oriental," "Negro blood," Hitler's "Aryanism," "the peculiar genius of America," "the logical Frenchman," "the passionate Latin"--all represent a belief in essence. A mysterious mana (for good or ill) resides in a group, all of its members partaking thereof. (Gordon Allport, *The Nature of Prejudice*, 1954: 173-174.)

On the horizon are tens of thousands of severely morally impoverished juvenile super-predators. They are perfectly capable of committing the most heinous acts of physical violence for the most trivial reasons... They fear neither the stigma of arrest nor the pain of imprisonment. They live by the meanest code of the meanest streets, a code that reinforces rather than restrains their violent, hair-trigger mentality. In prison or out, the things that super-predators get by their criminal behavior -- sex, drugs, money -- are their own immediate rewards. Nothing else matters to them. So for as long as their youthful energies hold out, they will do what comes "naturally": murder, rape, rob, assault, burglarize, deal deadly drugs, and get high. (John Dilulio, "The Coming of the Super-Predators," 1995)

We also have to have an organized effort against gangs. Just as in a previous generation we had an organized effort against the mob. We need to take these people on. They are often connected to big drug cartels, they are not just gangs of kids anymore. They are often the kinds of kids that are called "super-predators." No conscience, no empathy. We can talk about why they ended up that way, but first we have to bring them to heel. (Hillary Clinton, campaign speech in New Hampshire, 1996)

0. Introduction

Consider the epigraphs above. All concern the notion of *essentialism*-- the idea that all or most members of a social group or category share, deep down, a fundamental essence which causally grounds their observable traits, dispositions, and behaviors. Gordon Allport, the pioneering psychologist who is largely responsible for putting racism on the agenda for the psychological sciences, claims that a belief in national or racial essences is on the one hand a valuable information-processing device which "saves us from the pains of dealing with [people] as

individuals.” On the other hand, it is the foundation upon which group stereotypes and virulent racism are psychologically based. John Dilulio, the conservative criminologist who, along with a generation of policy makers, saw urban crime as *the* domestic existential threat to American security, gave those policy makers its essentialized bogeyman, in the form of the “super-predator.” In Dilulio’s vivid depiction, the super-predator is a nihilistic hedonist, driven “by nature” to commit all manner of heinous crimes, and highly resistant to reformation or correction. And finally, Hillary Clinton, then first lady in President Bill Clinton’s administration, exemplified the 1990’s punitive attitude to (non-White) crime when she singled out these “super-predators” as particularly deserving of punishment.

“Essentialism” can thus be understood as i) a general psychological disposition we employ in making sense of the world, ii) as a way of explaining behavior, and iii) as something which underwrites attributions of responsibility, guilt, and fitness for punishment. The purpose of this chapter is twofold. First, it explores the ways in which essentialist beliefs are used as *explanations* for the behavior of members racialized social categories. I show that conceptions of essence are often invoked to answer questions such as-- what is it about *Muslims* that makes them *terrorists*? Why are *Blacks* more *criminal* than *Whites*? Why do *immigrants* take advantage of *native* resources? Second, this chapter explores how essentialist beliefs might be used in attributions of *responsibility* for behaviors and traits. To put the point crudely, I argue that we should expect the particular conception of essence that is invoked to do this explanatory work to shift according to *how much responsibility* an interpreter wants to attribute to members of these racial groups.

The positive hypothesis about conduct-standards is that racial group membership has an effect on the degree to which a group is seen as *responsible* for its conduct. I will suggest that if an

outgroup racialized group R is perceived to have a negative racial trait, that trait is likely to be seen as i) rooted in R's essence, and ii) as something members of R are especially morally responsible (blameworthy) for manifesting. Moreover, if an outgroup racialized group R is perceived to have a stereotyped positive group trait, that trait will be seen as i) rooted in R's essence and ii) as something R is less morally responsible (praiseworthy) for manifesting. These valences are reversed when R is an ingroup--negative ingroup behaviors are seen as having exculpations, and positive outgroup behaviors are seen as being especially praiseworthy. This hypothesis certainly shares a family resemblance with various "attribution errors" which have been discussed in the literature on intergroup cognition. Yet as we will see, neither the *ultimate attribution error* (Pettigrew 1979) nor the *supreme attribution error* (Leslie 2017) satisfactorily explains this dynamic. Consequently, I dub my own proposal the *moralistic attribution error*.

The structure of the chapter is as follows. Section 1 presents the central concept of "essentialist belief" in more detail, and argues that there are good a priori reasons to think that a variety of racialized categories are essentialized. Section 2 presents empirical evidence that shows that four commonly racialized categories-- *immigrants, terrorists, criminals, and welfare recipients*-- are often bound up with essentialist racial beliefs. These essentialist beliefs provide (to an interpreter) partial *explanations* for why members of these categories have the properties they have. Drawing on some of the literature on intergroup cognition, Section 3 attempts to understand how essentialist explanations work in intergroup contexts. We will discuss two extant frameworks for understanding how group dynamics shape essentialist explanation, and find that both are confronted by empirical and conceptual difficulties. These shortcomings provide conceptual space for a new hypothesis about how intergroup explanation and attribution works, which I present in Section 5.

Yet first, there is a problem to be solved. Section 4 directly addresses the seeming tension between the notion that the *essence* of a racialized category R explains why members of R engage in behavior B and the notion that members of R are *responsible* for engaging in B. I argue that interpreters shift to more “plastic” conceptions of essence in order to explain the causal roots of B, and that these more plastic essences permit interpreters to hold members of R responsible for their behaviors. I suggest that this shows that cognition of racialized categories is broadly “Humean” in the sense that both essentialist explanatory strategies and attributions of responsibility are largely driven by “the passions.” Once the ground for the view has been cleared in this way, Section 5 presents the moralistic attribution error in more detail, and provides evidence for the hypothesis.

At the end of this chapter, we will be in a position to understand not only the basic structure of the representations of the racialized categories we are interested in, but also how these representations can undergird race-specific shifts in conduct-standards. The goal of the Chapter 6 will be to show how racialized discourse can transmit these sorts of representations.

1. Essentialism and Racialized Categories

In the sense we are interested in here, *essentialist* beliefs represent members of a kind as sharing a non-obvious underlying nature or essence which “grounds” the enduring properties that make members of that kind the same or similar in salient ways (Gelman 2003). Minimally, essentialist theories are committed to four claims: there is some group essence the possession of which i) “constitutes the individual’s membership in the kind,” ii) “causally explains the typical properties of that kind,” iii) is highly resistant to change, development, or loss, and iv) is transmittable or replicable in offspring (Mallon 2016: 26, Keil 1989). For instance, someone who has an essentialist

theory of the category *tiger* is likely to believe that all tigers belong to this category by virtue of sharing something like *tigerhood*, that possession of tigerhood is what explains the fact that all tigers have stripes, that even severe modifications of a given tiger's appearance (e.g. eliminating its stripes, adding horns, subtracting its legs) are insufficient to erase its essential tigerhood, and that tigers beget other tigers by virtue of a transmission of tigerhood.

Beyond this basic formulation, however, there is a good deal of metatheoretical disagreement in psychology concerning how best to operationalize “essentialist” beliefs. Haslam, Rothschild, and Ernst (2000) suggest, with empirical support, that essentialist beliefs about a category C come in two basic varieties. The first variety tends to represent C as a *natural* as opposed to artificial category, and usually prompts inferences that C is a *discrete* category with relatively sharp boundaries, *immutable* and resistant to development, *historically stable*, and as possessing *necessary* characteristics. The second variety tends to represent C as a *reified* or *hypostatic* category as opposed to a fluid or abstract one, and tends to prompt inferences that C has an *inherence* that causes basic sameness among its members, that membership in C is *informative* in a way that justifies a large number of judgments about its members, that members of C possess a high degree of *uniformity*, and that C is *exclusive* in a way that precludes its members from belonging to other categories. Going forward, I will assume this basic bipolar picture of the nature of essentialist beliefs. The “nature/reification” spectrum gives us a more specific understanding of the respects in which social categories are conceived as entitative. As we will see, the racialized categories we will be interested in here can be essentialized in both of these ways, depending on the experimental and social context.

The mechanisms responsible for generating essentialist beliefs cross-cut a variety of ontological domains. While humans seem to be especially liable to form essentialist theories of

common “natural kinds” such as tigers, birds, and trees (Gelman 2003), they also tend to essentialize gendered (Cimpian et al 2012), racial (Hirschfeld 1998), and even novel, previously unencountered social kinds (Waxman 2010, Rhodes et al 2012). Which sorts of categories this mechanism essentializes are highly sensitive to cultural context. Rhodes and Gelman (2009) show that people raised in politically conservative communities are more likely to hold entity theories about race than people raised in liberal communities. Western children essentialize races but not occupations such as “lawyer” or “athlete,” while Indian children tend to essentialize class and social position, not race (Mahalingam 2003.)

Finally, evidence suggests that there are rich causal interconnections between essentialism, stereotyping, prejudice, and discrimination. In keeping with the orthodoxy observed in much social psychology, we here reserve the term “stereotype” for the specifically *cognitive* attitudes which represent general features of a category, “prejudice” for the *affective* or emotive attitudes directed toward category members, and “discrimination” for the *behavioral* manifestation of these attitudes (Fiske 1998). Essentialists about a category are far more likely to form and endorse stereotypes about that category, and are more likely to pay attention to stereotype-consistent information rather than stereotype-inconsistent information (Bastian and Haslam 2006/2007, Prentice and Miller 2006, Williams and Eberhardt 2006, Yzerbyt, Corneille, and Estrada 2001). Essentialism also often predicts the endorsement of prejudice and dehumanizing attitudes toward a category (Haslam, Rothschild, Ernst 2000/2002, Bastian and Haslam 2008, Rhodes and Mandalaywala 2017). It is thus very important to get clear on what sorts of kinds and properties are typically essentialized at a given social context, since these will be the kinds and properties most at risk for being targeted by the attitudes that cause social marginalization.

1.1 *Essentializing racialized categories*

There are at least two methodological problems that confront an attempt to demonstrate that the sorts of racialized social categories we have been discussing are essentialized. First, as we've seen, terms like "terrorist," "thug," "criminal," "immigrant," and "welfare recipient" pick out social categories that are *ambiguously* racial.⁷⁴ Interpreters don't always interpret these terms as referring to specifically racialized populations, and so it's difficult to be sure that their essentialist representations of race are causally involved in their beliefs about *terrorists*, *criminals*, and the like.

Second, while there is evidence that local, culturally specific mechanisms can affect what sorts of categories are essentialized in common sense conception (Mahalingam 2003), a number of researchers contend that the human *capacity* to essentialize natural and social categories is evolutionarily ancient, species-specific, and innate (Sperber 1994, Hirschfeld 1998, Mallon 2016). Accordingly, we should expect the psychological mechanisms responsible for generating essentialist representations to focus on reasonably cross-cultural or fundamental kinds. Yet the categories of *thug*, *criminal*, and *terrorist* seem rather fine-grained and culturally local. It thus requires some work to show that either the mechanisms of essentialism or the essentialist representations they generate concerning "deeper," more "fundamental" sorts of categories (like natural kinds), are causally responsible for the acquisition of essentialist representations about more shallow, specific categories.

In response to the first worry, even if it is true that terms for racialized categories are "ambiguously racial," we will soon see that the categories those terms denote are often perceived in a racial light. In response to the second worry, the cultural specificity of a category C need not

⁷⁴ The notion of "ambiguity" I have in mind is not a semantic one. It simply refers to the fact that the category sometimes appears racialized to an interpreter in specific ways, and sometimes does not.

preclude the possibility that C is commonly essentialized. For one thing, widespread thought about Cs might incorporate “ingredient representations” of more stable, cross-cultural categories. *Thug* and *welfare recipient* certainly seem to be culturally local categories, but the core properties that those categories are perceived to instantiate-- *violence, aggressiveness, brutality, laziness, stupidity*, etc.-- are longstanding features common to human societies ancient and modern. Moreover, the evolutionary “canalization” of essentializing mechanisms is compatible with newer categories becoming essentialized. For instance, one can grant that as a result of Western colonialism, the concept RACE *means* something different now than it did in the 1400. We might even say that the identity conditions on the racial categories themselves have changed. None of this shows, of course, that the new racial categories are *any less essentialized* than the old ones were in 1400. We may be using very different terms to denote the categories, but the essentializing mechanism is just cranking out nominally different representations. As David Livingstone Smith puts the point, “the form of ethnoracial thinking is innate, while its content is determined by cultural beliefs and ideologies” (Smith 2011: 200).⁷⁵

⁷⁵ This is in effect the same argument that Mallon (2016) presents in defense of the claim that the concept RACE is substantially cross-cultural and pre-modern, despite the fact that folk theories of *race* are highly sensitive to cultural context. This contrasts with the views of many humanistic social constructionists about race, who often hold that races (as well as the RACE concept) are products of historically recent and culturally specific practices, such as the expansion of European empires, the transatlantic slave trade, and both domestic and satellite colonial regimes (Taylor 2004, Painter 2010, Omi and Winant 2013.) For Mallon, race-thinking exists wherever *lineage essentialism* exists-- the belief that similar-looking human populations create similar-looking human populations by virtue of transmitting a group essence. Since there is a good deal of empirical and historical evidence that lineage essentialism with regard to somatically diverse human kinds is both ancient and culturally widespread, this is evidence that race-thinking is ancient and cross-cultural, even if folk semantic and folk metaphysical theories of RACE/*race* show a high degree of cultural and temporal variation. (An example of this cultural/temporal variation is the slogan that “race doesn’t travel”-- given variations in practices of racial identification, the self-same individual can be “White in Brazil, Black in the US,” or an Irish-American individual who counts as White now would not count as White in 1840’s America.) Mallon’s hypothesis is obviously compatible with the general thrust of the social constructionist point, since it leaves open the possibility that while RACE is an evolutionarily recent representation, the concept/category RACE/*race* did not take on particularly evaluative or hierarchical forms of significance until it became theorized and weaponized during the colonial era of European expansionism (Mallon 2016, Chapter 1).

2. Immigrants, Terrorists, and Welfare Recipients

In this section, I review evidence for the claim that three racialized categories-- immigrants, terrorists, and welfare recipients-- are often essentialized in common sense conception. Members of these groups are commonly believed to share a common a non-obvious entity or inherent disposition which is causally responsible for their socially negative behaviors. Once we've reviewed evidence for this hypothesis, we will turn to a more general discussion of the mechanisms that lead interpreters to attribute category members' behaviors to their shared essence.

Taking a cue from Gordon Allport, I will assume in what follows that there is a particularly close connection between essentialism and *stereotypes*. Stereotypes are representations which ascribe properties and features which are believed to be widely shared by members of a social group. Moreover, these stereotypes often "link specific attributes to the very essence of what people are" (Yzerbyt, Rogier, and Schadron 1997). Research on stereotypes thus provides a valuable source of indirect evidence about what sorts of social groups are essentialized, and how. This research suggests that there is a bidirectional relationship between stereotyping and essentialism. First, harboring stereotypes about a group predicts entitative thinking about that group, plausibly because stereotypes underwrite overgeneralization effects which foreground similarities and background differences within the stereotyped group (Hamilton and Sherman 1996.) Likewise, essentialism seems to increase the facility with which subjects form stereotypes (Yzerbyt, Corneille, and Estrada 2001).

Of course, this connection between stereotyping and essentialism is defeasible. While stereotypes about lawyers and police officers abound, for instance, few of the properties of lawyers and police officers, positive or negative, are attributed to the *essence* of these kinds. Consequently, in what follows we will need to appeal to supplementary evidence in order to show that stereotypes

about racialized categories are indeed bound up with essentialism.

2.1 *Immigrants*

Immigrants are often perceived as having values which are antithetical to those of the dominant culture, as manifesting socially detrimental traits and behaviors, as scapegoats for national problems, and as the recipients of limited resources that “rightfully” belong to natives. Yet not all immigrants are perceived in the same way. In the United States, some immigrant groups are considered models of dedication to the ethos of hard work and individualism, while others are considered socially maladapted, lazy, criminal, or otherwise suspicious.

Fiske et al (2002) propose a general model of stereotype content which predicts that stereotypes of social groups are organized around values for “competence” and “warmth.” A target group is represented as “competent” if it is perceived as having a high or competitive social status and is able to both set and achieve whatever its goals are. “Warmth” is a measure of the extent to which members of the group are perceived to intend either benefits or harm. “Generic” immigrant groups explicitly unidentified by race, ethnicity, national origin, etc. generally score low in both competence and warmth. Yet Lee and Fiske (2006) show that perceptions of ethnicity, national origin, and class play a role in determining the competence-warmth “score” of immigrant subgroups. For instance, Asian immigrants tend to be represented as highly competent due to their perceived “model minority” status, but low in warmth due to their perceived status as “perpetual outsiders” who never fully assimilate to the native culture’s way of life (Kim 1999.) Canadians and European-Americans are treated as above-average for both competence and warmth, and European immigrants score slightly lower on both dimensions. Finally, four immigrant categories-- *Latino*,

Mexican, farm worker, and undocumented-- received the lowest score in both competence and warmth (*Middle Eastern* immigrants were judged low in warmth, but high in competence.)⁷⁶ (Lee and Fiske 2006: 759) The fact that Latino/Mexican/undocumented/farm working immigrants share a “low competence-low warmth” score with that of “generic” immigrants is at least circumstantial evidence that these categories are commonly represented as “the” prototypical immigrant group in the United States.

Fiske’s competence/warmth model is a useful heuristic for determining how individuals represent the mental states (warmth) and skills or abilities (competence) of immigrant outgroups. There is also evidence that these stereotypes are causal effects of more primitive forms of entitative thinking about social groups. For instance, Bastian and Haslam (2008) show that when members of national ingroups think of *themselves* as defined by a group essence, they are more likely to think of immigrant outgroups as entitative as well. These experiments showed that White Australian subjects primed to think of their own national identity as “exclusive” rather than “inclusive” were more likely to think of immigrant outgroups as sharing inherent dispositions that grounded their stereotyped properties and accounted for their perceived similarity. These “exclusive Australian identity” subjects were also more likely to possess strongly negative attitudes toward Asian immigrants. Similar anti-immigrant attitudes are associated with nationalistic attitudes more generally, which represent one’s own national character as coherent, unified, and superior to others.

Finally, there is evidence that immigrant groups are widely thought to possess a fundamentally different nature than native groups. Leyens et al’s (2001, 2003) research on intergroup

⁷⁶ African immigrants also occupied this dimension of the scale, but were considered overall “warmer” and “more competent.” The authors hypothesize that this average reflects the mixed effects of the popular Western perception that African immigrants are high achievers, but that the continent that they emigrate from is a violent, politically unstable cultural backwater.

attitudes between nationally and regionally defined groups suggests that entitative thinking about the social identity of one's own regional-national ingroup leads to the "infra-humanization" of members of a regional-national outgroups. Specifically, these outgroups are judged as having fewer of the mental states that are treated as essentially "human." They find that while members of ingroups recognize that outgroup members experience "primary" animal/human emotions such as anger, surprise, fear, joy, and disgust, outgroups are also judged to be less capable of more sophisticated "secondary" emotions such as "admiration, sorrow, fondness, disillusionment, contempt," and "conceit" (Leyens et al 2001). Since the ability to experience these sorts of emotions is widely taken to be a mark of full humanity (Leyens et al 2003), outgroup members are regarded as just a bit less human than ingroup members. As a prototypical outgroup, immigrants are thus often represented as sharing less fully in the human essence. This hypothesis is of course compatible with the suggestion that immigrant outgroups are *dehumanized* as sub-human (Smith 2011), and receives independent support from sociolinguistic work demonstrating that immigrants are both depersonalized as statistical representations and compared to animals in implicit and explicit ways within public discourse (Santa Ana 1999, Khosravini 2010).

2.2 Terrorists

A direct attempt to determine whether the category *terrorist* is essentialized as a function of a racial mode of presentation is confronted by a serious difficulty. The empirical studies (as distinct from polls) that focus directly on how naive subjects think and feel about terrorism and terrorists often presuppose a conceptual framework in which the members of the terrorist category are *already* characterized as Middle Eastern Muslims. Moreover, these studies often take place in the

wake of historical events involving Muslim terrorists (e.g. Fischhoff et al 2003, Coryn, Beale, and Myers 2004, Small, Lerner, and Fischhoff 2006).

Of course, this is not necessarily a failing of such studies, since this lacuna is not easily remedied. Because “the enemy” in the global War on Terror “has been repeatedly defined by its religious identity,” subjects queried about their views on “terrorism” are highly likely to assume that Muslims, and particularly Muslims of Middle Eastern or African origin, are under discussion (Sides and Gross 2013.) It is thus difficult to gather data about how terrorism is conceived *independently* of the ethno-religious-racial mode of presentation that often accompanies it.⁷⁷

On the other hand, the perceived close connection connection between terrorism and Muslim identity means that research on attitudes about Muslims frequently uncovers representations that straddle the boundary between religious identity and political extremism. For instance, Whites tend to believe that *both* Muslims and Muslim-Americans are violent and untrustworthy, and are less hard-working and intelligent than Whites (Sides and Gross 2013). This suggests that Muslims are believed to share an ethno-religious essence which is not diluted with differences in cultural upbringings.

Evidence suggests a number of possible sources of essentialist thinking about Muslims. Newheiser et al (2009) finds that anti-Muslim *prejudice* significantly predicted whether Muslims are thought of as an entitative group. The fact that a bidirectional relationship obtains among *both* entitative thinking/prejudice and entitative thinking/stereotyping is somewhat surprising, since meta-analyses of the relationship between stereotyping and prejudice suggest the causal relation between the two is tenuous at best (Dovidio et al 1996, Fiske 1998). Essentialism thus seems to be

⁷⁷ From the point of view of my semantic theory of racialized terms, this suggests that Muslim/Middle Eastern identity is defaultly taken to be central to the property threshold for “terrorist.”

functionally connected to “cold” cognitive processes like stereotyping and “hot” reactive attitudes like emotion and prejudice, even if these “cold” and “hot” processes are not directly connected.

Finally, a qualitative source of evidence concerning representations of Muslim *terrorists* comes from the massive literature produced by theorists who seek to *psychologize* individuals and groups that commit terrorist acts.⁷⁸ Interestingly, as a result, we are in possession of a good deal of indirect evidence concerning whether *theorists*-- including international relations experts, security experts, organizational and social psychologists, and psychopathologists-- harbor essentialist beliefs which affect their investigation of their target phenomena. If it is the case that these theorists essentialize terrorists--if they commonly assume or argue that there is some entity that drives individuals to commit terrorist acts-- perhaps such representations should be understood as part of a common theory of terrorism which extends outside the circle of experts and into the broader social environment. Legal scholar Ali Khan criticizes much of this literature as “a coordinated effort on part of academics, scholars, think-tankers, journalists and others to create a profile of Muslim militants as essentialist terrorists who... are spiritually addicted to violence”: (Khan 2005)

The essentialist terrorist is a violent monster that the Highly Influential Terrorist Literature (HITLit) has successfully invented and made real and believable.⁷⁹ It is the new terrorist. It is dark and evil, part real and part phantom, part human and part animal, part man and part woman, part bearded and part veiled, part strategic and part crazy. A noted trait of this grotesque but cowardly creature is that it kills innocents. But this monster’s most defining characteristic is that it is driven to violence by its nature, compelled by an ingrained mental/psychological/cultural/religious formation. Its violence has little to do with any outward political or geopolitical grievances. (Khan 2005: 47)

⁷⁸ See Gill and Horgan (2012) for an overview of some of these hypotheses.

⁷⁹ Of course, the claim that “HITLit” has “invented” or “made real” this “new terrorist” may be overstated. If it is true, as I suspect, that the presuppositions of this literature should be understood as borrowed from extant culturally distributed essentialist beliefs about terrorism, then HITLit does not so much as “invent” this terrorist so much as it systematizes these beliefs into a cohesive theory.

Indeed, as Leslie (2017) points out, qualitative evidence of this essentialist construal of terrorists, and Muslim terrorists in particular, is not hard to find. Bernard Lewis, widely regarded as one of the most influential Western interpreters of the politics and culture of the Middle East, describes the “culture of Islam” as being deeply connected to both “dignity and courtesy” as well as “explosive rage and hatred”:

There is something in the religious culture of Islam which inspired, in even the humblest peasant or peddler, a dignity and a courtesy toward others never exceeded and rarely equaled in other civilizations. And yet, in moments of upheaval and disruption, when the deeper passions are stirred, this dignity and courtesy toward others can give way to an explosive mixture of rage and hatred with impels even the government of an ancient and civilized country – even the spokesman of a great spiritual and ethical religion – to espouse kidnapping and assassination, and try to find, in the life of their Prophet, approval and indeed precedent for such actions. (Lewis 1990)

In each case, what seems to be responsible for acts of terrorism is an intrinsic, quasi-natural predisposition towards violence. Perhaps that underlying mechanism is an “underlying mental/psychological/cultural/religious formation,” or perhaps a “religious culture” given to an “explosive mixture of rage and hatred” which legitimizes unspeakable acts. The point is simply that the cause of terrorism is frequently believed to be an *entity* which lurks in the soul of the terrorist actor himself, and simply requires the right triggering conditions to express that underlying nature.

2.3 Welfare recipients

Sociologists and historians often ascribe Americans’ professed distrust of “the welfare state” to the perception that welfare is a “European” institution diametrically opposed to typically American virtues like government non-interference, individualism, hard work, and merit-based entitlements. There is a grain of truth in all this. Yet in his landmark study on the impact of racial

stereotypes on welfare policy, Martin Gilens (1999) argues that despite their individualism, Americans don't necessarily dislike the welfare state *as such*. Despite the welfare-bashing tactics of US administrations from Nixon to Clinton, for several decades running a majority of Americans continue to believe that the government should do more to improve the lives of the poor, the homeless, and the elderly, as well as spend more money on education and social security. What Americans unequivocally *dislike* is the idea that the powers of the welfare state are being used to pad the pockets of the "undeserving poor" who rely on welfare rather than working. Welfare as such is not perceived as anti-meritocratic or anti-individualist, but it is so perceived when welfare *recipients* are deemed to be undeserving.

For Gilens, American ambivalence about welfare is attributable to the facts that i) Whites believe Blacks to be particularly undeserving types welfare recipients due to their perceived lack of commitment to the work ethic, and ii) that many Whites falsely believe that *most* welfare recipients are Black and even that *most* Blacks are welfare recipients. As a result, Americans' approval or disapproval of actually existing welfare policies is often affected by their racial attitudes:

The centuries-old stereotype of blacks as lazy remains credible for large numbers of white Americans. The stereotype grew out of, and was used to defend, slavery, and it has been perpetuated over the years by the continuing economic disparities between black and white Americans. In a culture in which economic failure is often attributed to lack of effort, blacks' economic problems themselves reinforce the stereotype of laziness....this problem has been exacerbated by the emergence of a highly visible black urban underclass that has exerted an inordinate influence over popular images of blacks, even though it constitutes only a small fraction of African-Americans. (Gilens 1999: 3)

In a survey of over 1,700 White respondents, Hurwitz, Peffley, and Sniderman (1997) found that a majority of respondents agreed that "most blacks" are aggressive (50%) and lacking in discipline (60%), and that significant proportions of Whites agreed that "most blacks" are irresponsible (20%)

and lazy (31%) (Hurwitz et al 1997: 35). The question for Gilens is whether these stereotypes are involved in common attitudes towards welfare.

In collecting data for this hypothesis, Gilens (1999) asked nearly 1,000 non-Black subjects how likely a female welfare recipient was to look for a job while on welfare, and how likely a welfare recipient was to have more children in order to increase her benefits. All subjects were exposed to only one of two conditions. In one condition, the welfare recipient was White, and in the other, she was Black. This gave the experimenters a measure of what stereotypes interpreters endorse about White and Black welfare recipients, respectively. Subjects who thought it was very likely that a given welfare recipient (White or Black) would not look for a job or would seek to have more children in order to increase her benefits were treated as highly stereotype-endorsing.

Predictably, high-stereotype scores were more common in the “Black” condition than the “White” condition, even when controlling for level of education, political ideology, commitment to individualism, family income, and a number of other factors. Yet what is especially relevant for our purposes is that anti-Black welfare stereotypes in particular were much more significantly correlated than anti-White welfare stereotypes with commitment to a variety of general anti-welfare positions. For instance, subjects who harbored anti-Black welfare stereotypes were far more likely to agree that welfare recipients *in general* were undeserving or lazy. This strongly suggests that the category of *welfare recipient* is already racialized as Black in the minds of many Americans.

To say that welfare recipients are *essentialized* would be to say that interpreters commonly take their category-typical properties to be rooted in a shared group essence. Again, the fact that Blacks are an often negatively stereotyped population is indirect evidence that such a notion of group essence is often in play. As Yzerbyt, Rocher, and Schadron (1997) point out, representations

of Blacks are generated by two conflicting “models.”

In the first model, African-Americans are seen as being responsible for their current lot. An absence of motivation and proper values stands as the major reason for their being unable to achieve success in society. The second model states that structural disadvantages, that is, lack of job opportunities, inadequate education system, and so forth result in economic failure and lower social status. Whereas the first model characterizes African-Americans as the perpetrators of racial problems, the second model sees them as victims of discrimination.... The “perpetrator view” emphasizes the intellectual and motivational limitations of African-American people and so questions the impact of better school environments in critical neighborhoods. In sharp contrast, the “victim” view insists on the objective potential of African-American people and stresses the shortcomings of public investment in the school system. (Yzerbyt, Rocher, and Schadron 1997: 27)

To the extent that Black welfare recipients are perceived through the lens of the “perpetrator view,” welfare dependency will be linked to assumptions concerning the “intellectual and motivational limitations” of Blacks. The co-occurrence of these clusters of properties is then understood as undergirded by a deep, perhaps incorrigible disposition which provides reason to “question the impact” of things like “better school environments” for these populations.

2.4 Racial essences

To the extent that social categories like *immigrant*, *terrorist*, and *welfare recipient* are racialized, members of these categories are perceived to share a group essence partially rooted in racial features. We now need to address two questions. First, what role do essentialist beliefs play in the *explanation* of the behavior of members of racialized social categories? Second, what role do essentialist beliefs play in the attribution of *responsibility* for behaviors perceived to be characteristic of these categories?

3. Essence, Explanation, and Intergroup Attribution

When it comes to the role of essentialism in the explanation of behavior, what we want to

know is whether essentialist explanations are *selectively* invoked to explain the behavior of particularly racialized category members vis a vis non-racialized (or ingroup racialized) category members.

Likewise, when it comes to the role of essentialism in the attribution of responsibility, we want to know whether responsibility is *selectively* attributed to especially essentialized groups. If we can show that such explanations and attributions are selective, this would constitute good empirical evidence for the claim that race-related shifts in conduct-standards occur.

Let's put the hypothesis (too) simply. Consider two candidate descriptions of reality: *Blacks are criminal* and *Whites are criminal*. Assume that an interpreter thinks both of these descriptions is accurate. My claim about essentialist explanation is that an interpreter living under the conditions of racial domination is more likely to think that the truth of *Blacks are criminal* is explained by something about Blacks themselves, while the truth of *Whites are criminal* is explained by something about the situations that Whites find themselves in. My claim about responsibility attribution is that the interpreter is more likely to think that Blacks are *particularly* responsible for their criminality, while Whites are less responsible for having the same trait. As we will see, it will take some work to show that each of these hypotheses is consistent with the other.

In this section, we focus on the question of essentialist explanation. I begin by appealing to some relevant hypotheses about how behavior is explained in intergroup contexts. The evidence for these hypotheses gives us good reason to think that humans are more likely to essentialize members of social outgroups than they are their own ingroups. Consequently, we should expect racialized outgroup status to determine when essentialist explanations for the outgroup's behavior will be invoked. In particular, negative behavior perpetrated by members of racialized outgroups is typically thought to be a causal product of stable dispositional traits, or an inherent "character" rooted in a

group essence. In a parallel fashion, the negative behavior of ingroup members is often treated as a result of situational pressures. On the other hand, we will see that this literature has comparatively less to say about how essentialist explanatory strategies interact with intergroup attributions of *responsibility* for group-typical behaviors. This will bring us to a core tension between the notion of essentialist explanation on the one hand and the notion of responsibility attribution on the other. Sections 4 and 5 will be dedicated to resolving this tension and providing a unified account of how explanation and attribution work in the case of racialized categories.

3.1 A brief history of attribution errors

A longstanding critique of personality psychology (based in experiments such as Milgram 1963, Nisbett and Ross 1980, and formulated as philosophical critique by Harman 1999, Doris 2002) holds that character traits are insufficient to explain variations in human behavior, and are even poor predictors for future human behavior. More to the point, thinking of people *as* having these stable dispositions could be a manifestation of a more primitive cognitive bias known in psychology as the “fundamental attribution error” (FAE). Generally, the FAE causes an individual to attribute a stable dispositional trait in explaining another’s behavior, and to comparatively overestimate the role this trait plays in the production of human behavior. By the same token, the FAE leads individuals to underestimate the role that situations or environmental triggers play in causing behavior. As a result, people think that the underlying causes of others’ patterned behaviors are mostly due to features “internal” to those agents (e.g. motivations, intentions, character traits) rather than to “external” features of the environment in which those agents are embedded.

Sometimes the FAE manifests itself in the form of an “actor-observer asymmetry,” particularly in judgments about the causes of *negative* behaviors (Jones and Nisbett 1971). Individuals tend to explain the negative behavior of others by appeal to their robust character traits, while explaining their own negative behavior in terms of situational factors. For example, if Kendra cheats on her partner, Joelle might explain Kendra’s behavior by reference to her overall bad character-- she cheated because she is *dishonest, selfish, greedy, callous*, and so on. However, if Joelle herself cheats on her partner, she is likely to have a more charitable reading of her behavior, even if she recognizes that she did a bad thing by cheating. In particular, she will grant that she did something awful, but will appeal to extenuating circumstances which explain why she cheated-- perhaps her partner is being distant, or she shoulders a disproportionate share of the physical and emotional labor in her relationship, etc. This in turn generates asymmetrical moral judgments about whether these type-identical behaviors are justifiable. Although Kendra and Joelle have participated in the same highly negative behavior, from Joelle’s perspective, Kendra’s action is properly subject to greater disapproval, since it is the product of bad character.

What we are interested in is whether the tendency to explain others’ behavior by reference to stable dispositions and one’s own negative behavior by reference to situational pressures is also characteristic of *inter-group* behavioral explanations. Consequently, taking inspiration from the “fundamental” attribution error, a number of psychologists (Pettigrew 1979, Hewstone 1990) have sought evidence for an “ultimate attribution error” (UAE) which characterizes how group membership affects explanation of ingroup and outgroup behaviors. According to the UAE, members of social groups believe that negatively valenced behaviors by outgroup members are caused by a *disposition* toward the negative behavior, while similar ingroup members are believed to

be attributable to situational factors. By the same token, positively valenced behaviors by outgroup members would be attributable to situational factors, and similar behaviors by ingroup members would be attributable to stable dispositions.

There is scattered evidence for the UAE. Let's briefly consider a few representative findings. In a classic experiment in the context of the historically fraught relationship between Hindus and Muslims in Southern India, Taylor and Jaggi (1974) found that Hindu subjects were more likely to explain positive ingroup behaviors by reference to "internal" motivational features. Conversely, they were more likely to explain negative outgroup (Muslim) behaviors by reference to internal features. Similarly, Hewstone and Jaspers (1982) found that Whites and (West Indian) Blacks in the United Kingdom use different explanatory strategies in accounting for disparities in Black and White rates of arrest, unemployment, educational achievement, and occupational status. Subjects were asked whether the cause of Black disadvantage was due more to racial discrimination or to personal shortcomings. Whites attributed this disparity more to personal shortcomings on the part of Blacks, while Blacks were more likely to emphasize the role of racial discrimination. In Chapter 1, we mentioned the experiments in Duncan (1976), which found that White subjects were more likely to interpret an ambiguous action (a shove) as aggressive when the person doing the shoving was Black. Moreover, even when subjects recognized that White-initiated shoves were aggressive, they were more likely to attribute Black-initiated shoves to dispositional traits and White-initiated shoves to situational features.

Finally, and perhaps most relevantly for our current interests, Maass et al (1989) found that group membership affects what sorts of *linguistic* descriptions are used to characterize the conduct of ingroup and outgroup behaviors. The authors found that negative outgroup behavior was mentally

and linguistically represented at a “higher level of abstraction” than negative ingroup behavior. On the other hand, negative ingroup behavior and positive outgroup behavior was represented in more “particular” ways. Since abstract information is more difficult to disconfirm than specific information, examples of positive outgroup behavior are not treated as disconfirming. In a parallel fashion, negative ingroup behavior is easily disconfirmed, due to its particular, context-specific nature:

It is considerably easier to confirm or disconfirm the occurrence of a concrete behavior than the existence of a trait or psychological state. For instance, a single observation should be sufficient to disconfirm a false concrete statement such as "A hits B;" whereas many behavioral instances should be required before an abstract statement such as "A is aggressive" can be disconfirmed. The lack of verifiability should then make abstract statements more resistant to disconfirmation and change. (Maass et al 1989: 990)

One common way of encoding negative outgroup/positive ingroup explanations for behavior at a “higher level of abstraction” than negative ingroup/positive outgroup behaviors is describing the former in “dispositional” terms and the latter in “situational terms.” The authors found that subjects were consistently more likely to use dispositional descriptions in describing the negative behavior of outgroups, and more specific, situation-sensitive sorts of vocabulary in describing the negative behavior of ingroups, with the effect reversed for positive behaviors.

The findings we have just reviewed are highly suggestive. However, the evidence for the “ultimate attribution error” is confronted by a number of problems. First is simply the fact that the data is inconclusive: while there are many well-known documentations of the hypothesized effect, even these data provide only weak evidence for the claim that group identity *systematically* mediates behavioral explanations (Hewstone 1990). Moreover, the UAE cannot easily explain why members of social ingroups often endorse dispositional explanations for ingroup conduct and situational

explanations for outgroup conduct. In fact, ingroup/outgroup membership seems to be neither necessary nor sufficient for causing the effects hypothesized by the UAE. While group membership might get us started on thinking about why ingroup negative and outgroup positive behaviors are “explained away” by situations and why outgroup negative and ingroup positive behaviors are attributed to stable dispositions, it is clear that ingroup/outgroup membership must interact with other psychological mechanisms in order to cause the attributional patterns we observe.

3.2 *Essentialism and attribution*

There is good reason to think that *essentialism* mediates intergroup explanations of behavior. Positing essentialism as the “missing ingredient” in intergroup cognition is motivated by the fact that in many corners of developmental psychology, dispositional, agent-based explanations of behavior are often treated as expressions of underlying essentialist beliefs.⁸⁰ Moreover, there is a bidirectional relationship between essentialism and dispositional attributions: essentialist thinking about a group prompts dispositional attributions in explaining the conduct of group members. Yzerbyt, Rogier, and Fiske (1998) found that when a given outgroup was perceived as particularly “entitative,” subjects attributed group-typical behaviors to stable dispositions anchored in “underlying features [which] characterize the group,” and downplayed the role of situations in producing these behaviors. When the group was not conceived as entitative, there was no evidence of this tendency to attribute behavior to stable dispositions. Importantly, these results were gathered even though the “entitative” groups in question were highly artificial-- these groups were diverse collections of students who

⁸⁰ This connection is made explicitly by Gelman 2003, Cimpian and Markman 2010, Waxman 2010, Rhodes, Leslie, and Twarek 2012, and a wide variety of other studies. To be clear, in this chapter, I am treating dispositional attributions as defeasible but highly suggestive evidence that essentialist explanatory strategies are in play.

played a certain role within the context of the experiments, and took on “entitativity” as a function of playing that role, not as a function of having certain race, gender, class, etc. characteristics.

Accordingly, the study presents a strong example of the impact of entitative thinking on behavioral attributions.

Consequently, in addressing the shortcomings of the UAE, Sarah-Jane Leslie (2017) provides an independent argument for the idea that essentialism predicts what sorts of explanatory strategies will be invoked in intergroup contexts. Drawing on her work on striking property generics (Type B generics), Leslie introduces the “Supreme Attribution Error”:

The attribution error I have in mind here-- perhaps we should call it the Supreme Attribution Error, so as not to be outdone-- is concerned with a very different contrast [than that of the Ultimate Attribution Error.] We may very well explain negative behavior of in-group members by reference to their persistent personality traits, *but we will not explain it by reference to the group essence*. In contrast, when faced with highly negative behavior on the part of members of unfamiliar essentialized groups, we may take the disposition to such behavior to belong not only to the personality of the perpetrator, but to the very nature or essence of the group. (Leslie 2017)

Leslie openly admits that this “Supreme Attribution Error” (SAE) is in need of further development and evidence. In fact, she mentions that with the exception of the Yzerbyt, Rogier, and Fiske (1998) study mentioned above, “this particular attribution error has not been tested, nor even discussed” (Leslie 2017, *n* 43). Correspondingly, at this stage, we will evaluate the explanatory power of the hypothesis rather than its empirical support.

First, it is clear that the Supreme Attribution Error has a number of advantages. For instance, it nicely explains why interpreters often make strongly negative dispositional attributions for outgroup members, while *also* making parallel judgments about members of their own social ingroup. To take Leslie’s own example, for instance, White people often will admit that Osama

bin-Laden and Timothy McVeigh were unequivocally violent terrorists, and will grant that McVeigh's actions were the product of an evil and twisted character similar to bin-Laden's. We should not expect the fact that these Whites and McVeigh share a racial ingroup to make their judgments of McVeigh's conduct any less morally harsh.

On the other hand, Whites *are* much more likely to accept a generalization such as "Muslims are terrorists" than a generalization like "Whites are terrorists." As mentioned in Chapter 2, Leslie takes this distinction to be attributable to a difference in the way the psychologically primitive generalization mechanism functions with respect to different kinds. Plausibly, this generalization mechanism represents Muslims as a particularly *essentialized* kind, and thus the connection between being Muslim and being a terrorist is conceived as particularly non-accidental or informative. This is independent evidence that while terrorism of all sorts can be attributed to dispositional causes, Muslim terrorism is often represented as rooted in an underlying ethno-religious essence as well.

Consider one more example of SAE's explanatory power. As we saw, Americans think of welfare recipients as *in general* lazy and unmotivated, regardless of race. Leslie's hypothesis allows us to explain that while welfare recipients are in general thought of as lazy and unmotivated, Black welfare recipients are thought of as lazy and unmotivated *because* of their racial group membership. In other words, White and Black welfare recipients can both have a disposition toward laziness, but Black laziness is rooted in a deeper entitative cause. If the same sort of attribution mechanism is at work in the case of criminals, terrorists, and immigrants, this could fund a straightforward explanation of why different sorts of mechanisms are thought to explain the conduct of different racial groups.

Despite these benefits, however, there are two major shortcomings of Leslie's account. First, Leslie's SAE is explicitly designed to explain why *strongly negative* behaviors are taken to be rooted in group essences. As such, it is not supposed to explain why *positive* or *neutral* properties are often attributed to racial outgroups and ascribed to the group essence.⁸¹ Yet even stronger than this, Leslie claims that *we need not* explain cases involving generalizations about positive or neutral properties, since these simply aren't generalizations that subjects are prone to make.

It bears emphasizing once again that similar conclusions are not drawn when we are faced with neutral or positive information. This bias is not traceable to the essentialization of these kinds alone. If we look out of our window one morning and see a group of Muslims jumping on pogo sticks, we are not inclined to conclude something very general about Muslims. We do not suppose that the essence of the group grounds the disposition to jump on pogo sticks. When we encounter Muslim doctors, we do not conclude Muslims are doctors, and suppose that the non-doctors among them are nursing a yet-unmanifested disposition to practice medicine. And so on with an open-ended range of examples. The disposition to generalize in this way is specific to the strikingly negative, to the threatening, the vile, and the dangerous. (Leslie 2017)

Yet this seems puzzling. For one thing, a number of positively valenced generics about racial kinds are accepted, including "Blacks are athletic," "Asians are good at math," and "Muslims are disciplined."⁸² Moreover recall that Shelby (2003) points out that even very harmful racial ideologies seldom emphasize the racial roots of only negative traits. For instance, common stereotypes of Blacks are split between positive and negatively valenced features, and both of these are often thought to be expressive of the "Black essence":

The view that blacks are inherently of low intelligence is part of a wide-ranging and interconnected set of beliefs that includes, on the one hand, beliefs about the laziness, aggressiveness, and unreliability of black people, and on the other, beliefs about their

⁸¹ Thanks to Bernhard Nickel for pointing this out (Nickel ms.)

⁸² As we will discuss in Chapter 6, there is good reason to think that these generics are interpreted as saying something about how Blacks, Asians, and Muslims fundamentally are by nature.

natural musicality, athletic talent, and sexual prowess (Shelby 2003: 159).

While Leslie may be correct that attributions of striking properties are *more likely* than non-striking or positive properties to prompt hasty generalizations, it's puzzling why Leslie insists that positive properties aren't often treated as rooted in racial essences. Ideally, it would be preferable to have a uniform account of how essentialism is bound up in explanations for *both* positive and negative traits.

The second problem that faces Leslie's hypothesis is posed by the concept of *reduced attribution*.⁸³ Here's the general idea. Strongly negative attitudes towards Cs are triggered when those Cs are essentialized and racialized in particular ways. These sorts of attitudes often ascribe *moral failings* to Cs, and represent their behavior as morally repugnant. Yet if these behaviors are causal effects of an unchangeable, natural group disposition or essence, how could they be subject to moral assessment at all?

Consider an example. Tyrone, a Black man, and Will, a White man, each separately rob a convenience store. Bob, a White racist, condemns both Tyrone and Will, and insists that their behavior is rooted in dispositional traits like *greed, aggressiveness, stupidity*, and the like. Yet because of his anti-Black stereotypes and prejudices, deep down, he reserves more retributive attitudes toward Tyrone and directs less retributive attitudes toward Will (Hurwitz et al 1997). Assuming Leslie's hypothesis is correct, Bob thinks that both Tyrone and Will show bad character, but Tyrone's bad character is rooted in an underlying Black *essence* that makes him more disposed toward criminality.

If this picture is correct, it seems that Bob should give *Tyrone*, not Will, the moral benefit of the doubt. After all, Will's bad behavior is rooted in a mutable character trait which could respond to

⁸³ I borrow the label from Mallon (2016).

discipline and correction. But Tyrone's bad character is rooted in a deep-seated racial essence that makes him especially prone to criminality, aggressiveness, and violence. Unlike Will, whose life choices have played some role in cultivating his character, Tyrone has this essence solely by virtue of inheriting it. Assuming that this essence actually entails that Tyrone has *less* control over his criminal behavior than does Will, on what grounds can he be held (more) responsible for the actions that his essence predisposes him to commit?

Note that this isn't only a problem for Leslie. In fact, as Hewstone (1990) points out, reconciling the explanation of behavior with the attribution of responsibility is generally not a core focus of intergroup cognition research: "the focus on causal attribution deliberately excludes research on trait attribution (which does not address explanation) and responsibility attribution (where judgments are influenced by additional concerns such as moral evaluation and accountability)" (Hewstone 1990: 313).

However, our current concerns do not permit us to set aside questions concerning the attribution of responsibility across racialized intergroup lines. After all, part of the goal of a theory of racialized discourse is to explain the mechanisms that underlie race-related shifts in conduct-standards, and such conduct-standards inevitably involve attributions of responsibility, praiseworthiness, and blame. Consequently, the phenomenon of reduced attribution challenges us quite generally to show that essentialistic thinking about a category is compatible with moralistic thinking about that category.

The rest of the chapter runs as follows. Section 4 will show that moralistic thinking is both logically and psychologically compatible with *some* forms of essentialist thinking. Finally, Section 5

will sketch a general account of how racial presentation affects both explanations for behavior as well as attributions of moral responsibility for behavior.

4. Is Essentialist Thinking Compatible with Moralistic Thinking?

Again, the core assumption at the heart of the puzzle over essentialist and moralistic thinking is what Mallon (2016) labels *reduced attribution*. Here the idea is that when members of a racialized category R are represented as sharing a heritable, natural essence, judgments of moral responsibility for R-typical behaviors are *reduced*. When members of R are *not* represented as sharing a heritable, natural essence, judgments of moral responsibility for R-typical behaviors are *increased*.⁸⁴ The argument for reduced attribution depends on two further assumptions. First, attributors generally endorse a version of the well-known principle “ought implies can.” It is only rational to hold individuals responsible for their behavior if they have a certain degree of control over that behavior. Call this *control-dependent attribution*. And second, R-specific essences and the behaviors they cause are in general not the sorts of things that members of R can control. Call this *essence-uncontrollability*.

Consider a simple argument which we will dub *Reduced Attribution*:

Reduced Attribution

- P1. C’s group essence E is causally responsible for behavior B (essentialism)
- P2. Cs cannot control having E (essence-uncontrollability)
- P3. Cs are responsible for B to the extent they can control B (control-dependent attribution)
- P4. Cs have no/a low degree of control over B (1, 2)
- C. Cs have less responsibility for B than they would have in the absence of E (reduced attribution)

⁸⁴ This is slightly different from Mallon’s own formulation, which depends on the idea of “naturalness” rather than “essence.” See Mallon (2016): 95.

There is some empirical evidence that subjects endorse Reduced Attribution for at least some behaviors. Mallon (2016) summarizes a number of psychological findings:

John Monahan and Gloria Hood (1976) described a violent act as the product of a psychiatric disorder, and found significantly reduced judgments of both voluntariness and blameworthiness. Similarly, Bernard Weiner, Raymond Perry, and Jamie Magnussen (1988) found that describing a trait like *blindness* or a behavioral pattern like *being a child abuser* as the product of accidental or uncontrollable origins reduced judgments of culpability for the condition as compared to an origin under greater perceived control. More recently, John Monterosso, Edward Royzman, and Barry Schwartz (2005) found that describing behaviors as resulting from physiological causes (neurotransmitters) significantly lowered culpability assignments across a range of vignettes when compared to causes rooted in experience (e.g. abused as a child.) (Mallon 2016: 105)

If belief in both reduced attribution and control-dependent attribution are widespread, why are essentialized categories treated as particularly deserving targets of retributive attitudes? To return to our example, what explains Bob's ability to *both* think that Tyrone is essentially predisposed towards criminality *and* that Tyrone is an especially deserving target of retributive attitudes?

4.1 *Shifty Essences and Essentialist Explanation*

Perhaps the first thing to note is that the validity of Reduced Attribution depends on the causal role that "essences" are perceived to play in the causation of behavior. If the causal link between a category's essence and category-specific behavior is the sort of thing that is perceived to be under direct control, we should expect category members to be held responsible for such behaviors, and retributive or punitive attitudes will be indicators of when they are being held responsible. Yet we should *not* expect this outcome where the causal link is not perceived to be under direct control.

The second thing to note (recall) is that essentialist attitudes can be conceptually distinguished. Some essentializing attitudes represent a category C as *natural*, and others represent C as *entitative*. As Haslam, Rothschild, and Ernst (2000) point out, “entitative categories need not be naturalized, and categories understood as natural kinds need not be perceived to be entitative, but both are in some sense essentialized. In short, social categories may be essentialized in two distinct ways, and social psychologists should be heedful of both naturalness and reification/entitativity.” (Haslam, Rothschild, and Ernst 2000: 123). Consequently, in determining whether Reduced Attribution is true, we should be aware of which type of “essence” is understood to determine category-specific behaviors.

Representing a category as *natural* tends to reduce attributions of responsibility for category-typical behaviors. Thus, if the connection between racialized category membership and category-specific traits is generally perceived to be natural, then it’s unlikely that essentialist thinking alone underlies attributions of responsibility. On the other hand, there is evidence that *entitative* thinking about a category does not always reduce attribution of responsibility for category specific traits. Haslam and Levy (2006) found that those who regarded homosexuality as a biological/natural phenomenon were less likely to harbor anti-gay prejudice. This is very much in keeping with the evidence on reduced attribution-- if sexual orientation is beyond individuals’ control, there’s no sense in blaming them for having that orientation. However, Haslam et al (2002) found that *entitative* thinking about gays was significantly correlated with anti-gay prejudice. This suggests that even when homosexuality is considered controllable, or as a matter of personal choice, it can be treated as indicative as a fundamentally different underlying nature. In accounting for these results, Leslie (2017) suggests:

...[People] who believe that sexual orientation is a simple biological phenomenon do not consider gay people to have fundamentally different natures from heterosexual people....People who view homosexuality as a choice – and more specifically as a morally reprehensible choice – may take this decision to act immorally (by their lights) as indicative of a fundamentally different nature. (E.g. “I could never choose to do something so despicable, so anyone capable of making that choice must be deeply different from me and mine.”) Thus belief in shared nature can come apart from beliefs about biological bases, as this case perfectly illustrates. (Leslie 2017, n 35)

In the taxonomy of essentialist beliefs we have been working with, this suggests that individuals who regard homosexuality as a morally reprehensible choice conceive of homosexuals as sharing some underlying non-natural *entity* in common with one another. Yet attributing this entity does not reduce attribution of responsibility for homosexual behavior.

To summarize, if the relationship between membership in C and a trait T is perceived to be “natural” in character, T is seen as beyond C’s control. We should thus expect corresponding reductions in attribution of responsibility for having T. Yet *entitative* thinking about the relationship between C and a T need not represent T as immutable or uncontrollable. If T is represented as both rooted in a C-specific entity *and* under some degree of control from Cs, then entitative thinking will not reduce attributions of responsibility.

Now, what does this tell us about how racialized traits are conceived? Are Muslims terrorists by nature, and therefore not fully responsible for their actions? Is the perceived criminality, poverty, and indolence of Blacks rooted in a natural essence beyond their control? One of the things that prevents us from giving a decisive answer to this question is that interpreters often *shift between* “naturalistic” and “entitative” explanations in accounting for why a given racialized group has a given trait. For instance, in an experiment measuring views of intellectual ability across races, Ramsey et al (2001) recorded subjects’ racial views, and had them read one of two arguments. In the first

condition, subjects read an argument that located Black intellectual “inferiority” in bio-genetic causes. In the second condition, subjects read an argument that refuted the bio-genetic view. Strikingly, subjects with high degrees of anti-Black bias were happy to endorse *either* genetic or cultural explanations for Blacks’ perceived lack of intelligence relative to Whites, depending on which explanation was made most salient:

The results indicated that racists who read the counter-arguments adapted the cultural accounts to support their racist views. That is, those participants who were relatively high in racial hierarchicalization concluded that the reason for the putative inferiority of members of particular races rested in cultural deficits. (Condit et al 2004: 267)

That is, subjects who *would have* explained Black intellectual “inferiority” by reference to a natural bio-genetic essence ended up endorsing more entitative, “cultural” explanations for that same trait, solely based on which sort of explanation was salient at the time. This suggests that the question of when attribution for racialized traits is reduced is a highly contextual matter. Yet it also suggests that even traits like intelligence, which is often regarded as grounded by natural essences (Jayaratne et al 2009), is liable to be given an entitative reading when race comes into play. The authors suggest that this is a common feature of racist beliefs:

One attitudinal component that makes this surprising structure possible is the primacy of racial beliefs over causal accounts. That is, people do not first find a causal factor and then develop racist beliefs because they see that causation in operation. Instead, people first form hierarchical beliefs about races and then look around for explanations to account for those beliefs. (Condit et al 2004: 266)

We might think of this as a “Humean” picture of how thinking about racialized categories works. The “passions” (pre-existing prejudice towards racial groups, strongly negative emotions, etc.) play a role in determining what sorts of racial beliefs and explanations a subject will endorse on a given

occasion. Pairing this Humean picture with entitative thinking about race can have insidious consequences, because there does not seem to be a principled ontological limit on the “entity” that explains racialized behaviors. Whereas “natural” essences are usually taken to be innate and endogenous to group members (Haslam 2000, Gelman 2003), “entitative” essences are often highly relational properties like *culture*, *language*, and *religion*, which involve both group members and their environments. This suggests that group members can be held responsible for having a trait T even when the entity which explains the presence of the trait is highly relational.

To see this, consider another example of how pre-existing attitudes might cause an interpreter to switch between explanatory essentialist strategies. In a discussion of the “super-predator” myth which ascribes inherent criminality and predatoriness to young Black males, Tommy Curry (2017) argues that beliefs which represent Black criminality as natural or innate often co-exist with more “environmentalist” explanations of the causes of Black criminality. In one of the epigraphs that opens this chapter, criminologist John Diulio decries the environmental social conditions which have left urban Black men and boys “morally impoverished” while simultaneously emphasizing a supposedly innate, “natural” tendency towards destruction, hedonism, and predatory violence. One might think that these tendencies are explained by moral poverty, and that since morally impoverished agents have less control over their actions, that they should be held less responsible.⁸⁵ Yet according to Curry, super-predator mythmakers do not only see the external structure of “morally impoverished” environments as a competing *explanation* of violent crime. Rather, they also see these environments as a possible *symptom* of the deeper dispositions that ground

⁸⁵ This is of course is compatible with seeing some criminal actions as truly heinous. Yet we might take different responses to heinous crimes depending on what mechanisms we think underlie those crimes. Intuitively, environmentalist explanations might recommend policies that amend the environment, while entitative or naturalistic explanations might recommend policies that punish offenders.

poverty and crime. The result is “contradictory thinking about race and racism that asserts that the sociological conditions, the external environmental problems in poor Black urban spaces, is *evidence* of a more fundamental and inheritable cultural disposition toward deviance, criminality, and murder” (Curry 2017: 111-112, emphasis mine).

We can explain the “contradictory thinking” that Curry identifies by reference to our two conceptions of essence. The notion of “moral poverty” might be expected to shift our moral attention away from the individuals that engage in violent crime toward the structural conditions that create such individuals. But this has not happened in the case of the “super-predator” and attendant crime legislation. Why not? One reason is that policymakers understand “moral poverty” *both* as a complex of properties underwritten by the *natural* disposition of individual criminals *and* as itself the *cultural entity* that underwrites violent crime. “Moral poverty” is both “entitative” *explanans* and “natural” *explanandum*. Each “explanation,” of course, could justify punitive treatment: the “natural” violent dispositions of Black urban youth must be “brought to heel,” in Clinton’s words, in the pragmatic interests of public safety, and the cultural “entity” that is Black violence must be eradicated on more moralistic grounds. Either explanatory strategy can be accessed depending on what sort of treatment policy makers are especially interested in.

Thinking of essentialist explanation as roughly “Humean” in the sense we have outlined thus resolves the tension with Reduced Attribution. The shiftiness of essentialist thinking allows interpreters to emphasize whichever conception of essence seems to justify the sorts of *moral* or *reactive* attitudes they antecedently hold. In the next section, we consider the possibility that this represents a new kind of attribution error, and show that this insight helps us unpack the notion of a “race-specific shift in conduct-standards.”

5. Moralistic Attribution

Our brief review of the group attribution literature showed that neither the “ultimate attribution error” nor the “supreme attribution error” furnishes a satisfactory account of how essentialist thinking, responsibility attributions, and the reactive attitudes combine to produce differential judgments about similar traits and behaviors across racialized group lines. The problem with the UAE is simply that facts about group membership are neither necessary nor sufficient to explain these differential attributions. Leslie’s “supreme attribution error” (SAE), on the other hand, i) cannot explain the fact that *positive* or *neutral* properties are often regarded as explained by a racial outgroup’s essence, and ii) has very little to say about how essentialist explanations for behaviors can have effects on what sorts of moral responsibility attributions are made for those behaviors. As we saw, we needed to supplement the SAE by saying more about why responsibility for negative traits is not reduced even when those traits are seen as caused by a deep, underlying group essence.

My proposed moralistic attribution error (MAE) is an attempt to show how facts concerning essentialist explanations, responsibility attributions, and the reactive attitudes might co-vary. The core “Humean” hypothesis is that different explanations for the behavior of racial group R are accessed *depending on what sorts reactive attitudes concerning R an interpreter harbors*. The basic principle is that where R is a group targeted by generally antipathetic attitudes, interpreters seek to attribute *maximal* responsibility for R’s negative behaviors, and *minimal* responsibility for R’s positive behaviors. Attributing maximal responsibility allows an interpreter to exercise their reactive attitudes, since controllable negative and positive actions are especially good candidates for blame and praise,

respectively. Attributing minimal responsibility, on the other hand, reduces the appropriateness of blame and praise.

Recall that in Chapter 1 we gave one possible operationalization of the idea of a race-specific shift in conduct-standards. The basic idea was that an agent's racial membership might play a role in determining the relative acceptability of that agent's performance in context. We can now afford to be a bit more explicit. Race-related conduct-standards shift in at least three specific ways: i) with respect to the perceived *explanation* of conduct, ii) with respect to the perceived *degree of responsibility* for conduct, and iii) the *appropriateness of punishment* for conduct.

Race-related shifts in conduct-standards: A judgment J concerning an action A is reflective of a race-related shift in conduct standards between racialized groups R and R' just in case J reflects a shift in *Explanatory Strategy*, *Responsibility Attribution*, or *Punishment Estimation*:

a) *Explanatory Strategy*: The actions of members of R are often attributed to a shared group essence, which can either be "natural" or "entitative" in character. The connection between R's actions and group essence is represented as highly defeasible.

b) *Responsibility Attribution*: Perpetrator of group R, but not a perpetrator of group R' is deemed especially responsible for A when A is negative, and perpetrator R', but not perpetrator R, is deemed especially responsible for A when A is positive

b) *Punishment Estimation*: Perpetrator R, but not perpetrator R', is deserving of especially punitive responses when A is negative, and Perpetrator R', but not perpetrator R, is deserving of especially positive responses when A is positive.

Each of these shifts can occur separately or in tandem. What I am calling the *moralistic attribution error* (MAE) covers any case in which all three of these conditions is satisfied, given a fixed interpretation of R and R'. That is, the MAE covers any case in which a shift in explanatory strategies, responsibility attributions, or punishment estimations occurs "as a function of" at least one of the other two conditions.

I here remain non-committal about the contextual conditions that need to be satisfied in order to trigger the MAE. As we saw toward the beginning of this chapter, it is likely that widespread prejudice and stereotypes against racial groups, social-political tensions, and legacies of racial domination all play significant roles. What is important for my concerns here is that ingroup and outgroup membership play a far less central role in the MAE than they do in the broader group-attribution literature. My account makes room for the possibility that the individual whose judgments reflect race-specific shifts in conduct-standards may belong to either R, R', or some other group altogether. This means, among other things, that members of race R can commit the moralistic attribution error with respect to their own group.⁸⁶ My account thus treats ingroup or outgroup membership as a significantly weaker predictor for shifts in conduct-standards than does the UAE or the SAE.

However, since the MAE is merely a working hypothesis and not a fleshed-out theory, I will be unable to defend it in full detail here. In closing, I propose a few candidate examples of the phenomenon.

5.1 Muslim and non-Muslim "terrorism"

Terrorism is widely conceived as a particularly egregious sort of wrong, involving morally awful intentions to stir up fear for political gain, morally repugnant actions involving either mass murder, heinous abuse of individuals, or assault on the basic principles of freedom, and perpetrated in a spirit of both brutality and cowardice (Goodin 2006). Terrorists are conceived as some of the worst

⁸⁶ This makes the hypothesis *prima facie* compatible with a good deal of evidence from social dominance theory and system justification theory (discussed in Chapter 6). Much of this literature is motivated by the problem of explaining why members of subordinate groups often derogate members of their own group.

evil-doers, perhaps comparable to serial killers and child molesters. Calling someone a “terrorist” is thus a particular way of holding an individual responsible for their conduct, and is deeply bound up with the reactive attitudes.

As we have seen, Muslim terrorists are often believed to share a group essence which underlies violent behavior. This underlying Muslim/terrorist essentialism seems to be linked to corresponding shifts in both responsibility attribution and punishment estimation. In one series of studies, Small, Lerner, and Fischhoff (2006) found that dispositional attributions for “terrorist” behavior could be manipulated by priming subjects’ emotions in different ways. When primed to feel anger, subjects were much more likely to use dispositional and responsibility-attributing language in describing and explaining terrorist involvement in the September 11, 2001 attacks. Conversely, subjects were much more likely to use causal, situational language in describing these events when they were primed to feel sadness. Since dispositional attributions are often treated as a manifestation of underlying essentialist thinking about a category, this suggests that the reactive attitudes play a significant role in explanation and responsibility attribution.

Another application of the MAE here might be in accounting for why violence perpetrated by Muslims is frequently publicly classified as terrorism while violence perpetrated by Blacks, Whites, Asians, etc. is rarely so classified. This pattern persists even when the causes and consequences of violent actions are strikingly similar. In a review of domestic bombings, attempted bombings, and mass shootings in the post-9/11 era, Yin (2012) finds that the perpetrator’s identity is significantly predictive of whether they will be regarded as a “terrorist” or not by both media and law enforcement. In many cases, Muslim identity better predicts deployment of the “terrorist” label than does the intentions of the perpetrator, the “success” of the action, or the extent of the

destruction. For instance, consider the fact that none of the dozens of mass school shootings within the US over the past several years have been classified as terrorist attacks, and that the gunmen in such incidents are overwhelmingly White and male (upwards of 95%.)

One disparity which is of central interest from the perspective of the MAE is the extent to which White violence is treated as a symptom of an underlying aberrant “mental health” disposition and Muslim violence is treated as an expression of an underlying essential nature (Metzl 2017). Mercier et al (2018) show that subjects with significant anti-Muslim prejudice eliminate “mental illness” as a significant cause of mass shootings when presented with historical cases involving Muslim perpetrators. Such subjects were more likely to advance “mental illness” explanations in cases involving non-Muslim perpetrators, and even to conclude that a shooter described only as “mentally ill” was less likely to be Muslim.

Recall that in our discussion of Reduced Attribution we found that believing a behavior to be “naturally caused” often reduces attributions of responsibility for both “positive” and “negative” traits. One of the domains in which this effect is particularly robust is in attributions of responsibility for criminal acts caused by psychiatric disorders or uncontrollable neurological idiosyncrasies (Hood 1976, Weiner et al 1988, Monterosso et al 2005.) Where perpetrators are seen to have less control over the traits that cause their actions, they are often seen as less responsible for those actions. Believing that heinous acts committed by White shooters is attributable to mental illness might moderate responsibility attributions for those acts.

Believing that Muslims are driven toward terrorism by a group disposition and that non-Muslims are driven toward violence by situational features such as mental illness also has the effect of representing non-Muslim violence as *anomalous* or *aberrant*, and of representing Muslim

violence as in keeping with the general laws governing the Muslim “type.” Yet as we saw above, the perceived law-like connection between Muslim identity and terrorism need not be conceived in terms of a *natural* essence which reduced attribution-- in fact, we should expect the explanatory strategy to foreground “entitative” notions of essence.

It thus seems that the MAE is right to predict that essentialist explanatory strategies will often co-vary with corresponding shifts in responsibility attribution and in the strength of reactive attitudes. This coincides with Mercier et al’s judgments of the significance of their own results: “those with anti-Muslim prejudices are motivated to perceive Muslim mass shooters as less mentally ill, likely to maintain culpability and fit existing narratives about Islamic terrorism” (Mercier et al 2018: 1).

5.2 White and Black welfare recipients

The MAE might also shed some light on the connections between essentialism, responsibility attribution, and the reactive attitudes in the case of welfare recipients. Recall Gilens’ (1999) landmark experiments on racial attitudes and welfare policy. Americans attribute *more or less responsibility* to welfare recipients for their current lot, depending on whether those recipients are Black or White. We’ve already mentioned that Americans often explain the persistence of Black economic and social failures by reference to a racial essence which grounds negative traits such as laziness. Yet Americans also seem to think that Blacks are less *deserving* of assistance than Whites. One natural way of rendering this essentialist assumption and this “desert” assumption consistent is to hypothesize that Blacks are represented as having *more control* over the traits that cause their welfare dependency than Whites are. For instance, perhaps White welfare dependency is defaultly

explained by a perceived lack of control over the external environment. This might suggest that Blacks are seen as *more culpable* for their dependency than are Whites.

As it turns out, this reading of the data is corroborated by evidence that suggests that Americans harbor more punitive attitudes toward Black welfare recipients than to White recipients. In one survey of over 1,000 non-Blacks, researchers found that harboring anti-Black welfare stereotypes was between *two to four times* more predictive than anti-White welfare stereotypes of acceptance of claims such as “people are poor because they don’t try hard enough,” that “most people who don’t succeed in life are lazy,” that “most people on welfare could get by without it,” and most tellingly, “lazy welfare recipients make me angry” (Gilens 1999: 100). In other words, the poor are more culpable and more deserving of punishment when they are represented as non-White.

Similarly, Hurwitz et al (1997) find more direct evidence that those who harbor anti-Black stereotypes judge the Black poor more harshly when they are directly compared to the White poor:

When responding to scenarios of blacks who confirm (or only mildly disconfirm) their negative expectations-- welfare mothers, welfare recipients with poor work histories, or drug suspects using foul language-- [respondents who embrace negative racial stereotypes] consistently display a discriminatory double standard, such that they offer far more harsh judgments of blacks than of similarly described whites. (Hurwitz, Peffley, and Sniderman 1997: 52)

Consistent with our earlier discussion of the super-predator mythos, it is possible that these particularly stark attributions of moral failure are linked to an “entitative” essentialist explanatory strategy which sees Black failings as rooted in a constitutive “cultural” unwillingness to work their way out of poverty.

5.3 White and Black athleticism

So far we have been focusing on how race intersects with explanations and attributions surrounding generally negative traits. In closing, consider how the MAE might be applied to a positive, or at least neutral, trait: athleticism.

A number of studies indicate that a significant proportion of Whites believe that Black athletic ability is innate or biologically endowed, rather than acquired through motivation and training, while White athletic ability is acquired through grit and determination (Sheldon et al 2007, Harrison et al 2011). As such, a sort of “genetic essence” is deemed causally responsible for Black athletic prowess, while White athleticism is often attributed to more plastic or situational factors.

While “athleticism” is often regarded as a positive or neutral trait, this basic difference in explanatory strategies for Black and White athleticism is related to a network of negative stereotypes concerning Black abilities. Perhaps the most pervasive ur-stereotype in the vicinity is the “trade-off hypothesis”: the notion that Blacks have greater natural athletic abilities as a result of an evolutionary trade-off between bodily competence and “mental” talents. Yet the notion that Blacks are athletically gifted “by nature” can also be situated in the context of beliefs about Blacks’ (lack of) work ethic-- since Blacks are *ex hypothesi* incapable of achieving their ends due to laziness and lack of motivation, the explanation for those excellences they do display must be “natural” in origin. Consistent with the more general sorts of links between entity theories, prejudice, and stereotyping discussed in section 2, Sheldon et al (2007) found that the belief that a “genetic race difference” explained differences in Black and White athleticism “predicted [subjects’] level of prejudice toward and stereotyping of Blacks” (44).

It’s useful to contrast the common image of “natural” Black athleticism with that of White athleticism. White Americans often take Blacks to be more athletic than Whites, and that it is Blacks’

genetic endowment which gives them their athletic edge. Obviously this assumption can be understood in the context of the accompanying anti-Black stereotypes just mentioned, but it can also be understood against the backdrop of broader negative stereotypes concerning *White* abilities, including a constitutively arrhythmic nature, a lack of acquaintance with their own bodies, and so on. Yet interestingly, while Whites often believe Black athletic *promise* to be rooted in genetics, they do not believe White athletic *inferiority* to be rooted in genetics (Harrison et al 2011.) Whites overwhelmingly attribute White underrepresentation within grueling, kinetic sports such as basketball and football to cultural and socioeconomic causes.

From the point of view of the MAE, this seems to be a shift in explanatory strategies in the face of race. While Whites are content to describe Blacks as sharing a group essence that grounds observable traits, they are less likely to think of themselves as sharing such a natural essence. Yet the question remains: if Blacks' hypothesized superior athleticism is genetically endowed, why isn't Whites' hypothesized lack of athleticism genetic in origin? Why does natural endowment underwrite athletic talent in the case of Blacks, but not athletic limitations in the case of Whites?

From the point of view of the MAE, we have a ready proposal for why Black athleticism is explained by appeal to a natural predisposition. Consistent with the literature on Reduced Attribution, the perceived "naturalness" of a trait might affect judgments of how *responsible* the athlete is for her talent, and perhaps even how appropriate it is to praise her for possessing it. This is consistent with White views of Black athleticism: if Blacks possess a natural disposition toward athleticism, they exercise less control over its manifestation. Black athletes are thus less responsible (and praiseworthy) for possessing the trait.

What is more puzzling from the point of view of the MAE is why Whites do not explain

their own (lack of) athleticism in terms of genetic endowments. After all, that sort of naturalistic explanation would reduce attribution for White athletic *inferiority*, and increase attribution for those Whites who display athletic *competence* or *superiority*. The latter sort of “increased attribution” is very much on display in popular discourse concerning White athletes, which often emphasizes perseverance and moral strength. For instance, in sports media coverage, which by far represents the most extensive and wide-ranging body of evidence for the racialization of sports-relevant traits, White athletes are often portrayed as gritty, hard-working “grinders” who heroically succeed against the odds (Billings and Eastman 2002, Eastman and Billings 2001). So why *don't* Whites think of themselves as having a naturalistic, non-athletic essence?

One possible explanation that makes sense from the point of view of the MAE is that Whites do not perceive athleticism to have any non-accidental connection to whatever the *White* essence is (if any.) Appeals to White essences are just not treated as the sort of things that ground acceptable explanations of White abilities. This frame would thus represent Blacks as particularly *advantaged* by nature, while not representing Whites as particularly *disadvantaged* by nature. White athletic achievement would thus be cast as particularly meritorious success in the absence of a natural advantage, rather than as success in the presence of a natural disadvantage. Second, it may be the case that Whites *do* perceive White essence as relevant to athleticism, but that the essence in question is of the shifty, highly relational “entitative” sort discussed above. As we saw in the case of the perceived shortcomings of “Black culture,” this sort of essence should *not* be expected to reduce attribution.

5.4 MAE as a working hypothesis

Much more must be said in defense of the MAE hypothesis and the possible contexts in which it manifests itself. There are several possible connections to various corners of the psychological literatures on responsibility attribution, causal explanation, and motivated cognition that I have not explored here, and further engagement with these literatures would be a natural place to begin that work. Yet we have seen that the basic idea of the MAE represents a profitable intervention into three rather different social psychological literatures relevant to racialization-- explanation of behavior, responsibility attribution, and the reactive attitudes--and may even lend a degree of systematicity to these disparate findings. Moreover, it goes beyond its main competitors-- the “Universal Attribution Error” and the “Supreme Attribution Error”-- by reconciling widespread Essentialism with widespread Reduced Attribution. Finally, in the context of our theory of racialized discourse, the MAE gives us one way of characterizing the notion of a race-related shift in conduct-standards. This in turn gives us a more concrete grasp on what it could mean for individuals and groups to be held to different standards of conduct depending on their racial designation.

6. Conclusion

The goal of this chapter has been to give content to the idea that conduct-standards can shift systematically in the face of race, and to explore some of the psychological mechanisms which might explain how these shifts occur. We’ve described three such mechanisms. First, racialized social categories are likely to be *essentialized* in common sense conception, and the putative essence that the category shares is taken to be causally responsible for the traits and behaviors that members of these categories manifest. Second, members of racialized social categories are especially likely to be

regarded as *responsible* their socially negative or “pathological” behaviors. Finally, members of racialized social categories are deemed especially deserving targets of *punitive and retributive attitudes*.

We also saw that contrary to surface appearances, essentialist thinking about racialized groups is compatible with moralistic thinking about those groups. The chapter presented a hypothesis which reconciled essentialist explanation with moralistic attribution, and suggested that these mechanisms might be at work in many manifestations of race-specific shifts in conduct-standards.

So far our discussion has regarded the problems of accounting for racialized language and racialized thought in relative isolation. The goal of Chapter 6 will be to demonstrate the relevance of each domain to the other. There we will see that the production and interpretation of racialized discourse among speakers and hearers involved in everyday linguistic exchanges prompts precisely the essentialist explanatory strategies discussed above.

Chapter 6. Racialized Discourse and the Flow of Information

Abstract

We have proposed a semantic account of a certain region of racialized discourse and a psychological account of the ways in which racialized representations are deployed to explain and evaluate the conduct of members of racialized categories. The goal of the present chapter is to synthesize these accounts in a way that makes clear how racialized discourse is related to racialized thought. I will be concerned to show that racialized discourse both transmits essentialized representations of racialized groups and triggers essentialist explanatory strategies.

In arguing for this synthesis, I show that essentialist explanatory strategies are deeply bound up with interpreters' competence with generic language. Research in developmental and social psychology suggests that generic language is particularly likely to trigger the essentialist explanatory strategies discussed in Chapter 5. Since the examples of racialized discourse we have focused on here-- Type C generics, and a large subset of what I have called "racialized terms"-- are both species of generic language, we should expect each of them to have similar effects on the flow of information among interlocutors. Through an engagement with recent work in pragmatics and social psychology, I suggest that in addition to making salient essentialist explanatory strategies, these forms of racialized discourse play an indirect role in legitimizing social hierarchy.

“Society’s tendency is to maintain what has been. Rebellion is only an occasional reaction to suffering in human history; we have infinitely more instances of forbearance to exploitation, and submission to authority, than we have of revolt. Measure the number of peasant insurrections against the centuries of serfdom in Europe--the millennia of landlordism in the East; match the number of slave revolts in America with the record of those millions who went through their lifetimes of toil without outward protest. What we should be concerned about is not some natural tendency toward violent uprising, but rather the inclination of people, faced with an overwhelming environment, to submit to it.”

--Howard Zinn, *Disobedience and Democracy* (1968), pp. 16-17

“The only good Indian is a dead Indian.”

--Gen. Philip Henry Sheridan (attributed), c. 1869

0. Introduction

We have gradually been moving from a description of the semantic properties of racialized representations to a description of the potentially marginalizing effects of those representations. Chapter 5 suggested that the *linguistic* representations transmitted by racialized discourse can be bound up with the *cognitive* representations of racialized categories that characterize common sense conception. In so doing, we’ve made a case for thinking that the production of racialized discourse is a causal product of how racialized categories are represented. Yet we have also been assuming that the converse relationship holds as well-- that how racialized categories are represented in common sense conception is a causal product of racialized discourse. The work of the present chapter is to argue for this latter assumption.

Our goal will be to specify how racialized discourse both transmits and triggers racialized representations, and what sorts of pragmatic effects those representations have on the flow of information among interlocutors. I argue that racialized discourse has two such effects. First, racialized discourse directs conversational attention on *essentialist* explanations for negatively racialized traits such as criminality, aggression, and laziness. These explanations give special weight to

individual-based rather than structural explanations of these phenomena. Second, racialized discourse purports not only to *describe* racial realities, but also to *legitimize* those realities.

The core argument of the chapter is that both of these pragmatic effects are explained by the fact that the examples of racialized discourse we have been discussing are instances of generic language. Section 1 reviews recent evidence concerning the connections between generic acquisition and interpretation and essentialist thinking. In Section 2, I argue that understanding (some) racialized discourse as generically structured allows us to trace the sorts of effects racialized discourse has on the flow of information among interlocutors. Following the model of conversational context pioneered by Robert Stalnaker and recent work in update semantics, I suggest that generic language has two such effects on the possible worlds included in the context set (understood less formally as conversational “live options” during a given stage of the conversation.) First, generic sentences such as “Blacks are thugs” or even unembedded uses of “thug”⁸⁷ carry information that worlds in which Blacks are thugs or that thugs are generally Black are more *likely* than worlds in which they are not. Call this an *alethic* ordering. Second, these sentences carry information that worlds in which Blacks are thugs or that thugs are generally Black are *preferable to* worlds in which they are not.⁸⁸ Call this a *preferential* ordering. Jointly, these ordering mechanisms allow generic language to carry information about how the world *is* and also of how it would be good for the world to be.

⁸⁷ Recall the discussion of candid racial communication (CRC) in Chapter 3. There we found that unembedded lexical items can be and are given race-specific interpretations, even when such communication makes scant use of generics. CRC shows that speakers might use a variety of rhetorical tactics to prime speaker dispositions in a way that leads them to assign race-specific interpretations across a variety of conversational contexts.

⁸⁸ Obviously the *truth-conditions* of “Blacks are thugs” and “Thugs are Black” differ, and they would be deserving of different semantic analyses within the framework offered within Chapter 2. The thoughts expressed by these two sentences are similarly distinct, and have different accuracy conditions. However, here I assume that *either or both* of these representations might enter the common ground in both embedded and unembedded uses of “thug.” In the light of the classic psychological studies on humans’ rather poor performance in online reasoning with conditionals (Wason 1968) and probabilistic reasoning (Tversky and Kahneman 1983), I take this assumption to have a good deal of *prima facie* plausibility.

Section 3 is dedicated to working out the details of this (admittedly, ostensibly implausible) proposal. I begin with a discussion of Sally Haslanger’s provocative work on the pragmatic effects of generic language. Haslanger suggests that many social kind generics of the form *Fs are G* carry two sorts of implicatures into the common ground: that *Fs are G by nature*, and that *proper Fs are G*. While Haslanger’s story is rich and promising for a variety of social kind generics, I show that her theory does not adequately model the pragmatic profile of the types of racialized discourse we are interested in. Consequently, Section 4 concludes the chapter by proposing a new account of the function of racialized discourse which gives pride of place to the relationship between racialized discourse and *status quo* biases.

1. Generics and Essentialism

Recall our discussion of the “slippery truth-conditions” of generics in Chapter 2. Generics such as “birds fly” strike us as true, “books are paperbacks” strikes us as false, and “ducks lay eggs” strikes us as true. If how widespread a given property is among a given kind is relevant to the truth of a generic, this is a puzzling state of affairs. After all, almost all birds fly, the vast majority of books are paperbacks (circa 80%), and fewer than 50% of all ducks lay eggs.

One reason for this puzzle is that unlike kind-wide quantifications like “*all/most/some* birds fly,” generics express relationships between properties and *kinds*, not between properties and individuals. *How many* individual *K*’s are *F* is frequently irrelevant to the truth-conditions of ‘*K*’s are *F*’. Generics also express special kinds of relationships between kinds and properties (Prasada and Dillingham 2006). They do not simply imply a statistical, accidental, or chance correlation between a kind and a property; rather, they imply--and are commonly interpreted as implying-- a nomic, causal

connection between the two. “Ravens are black” does not mean that it just so happens that ravens tend to correlate with blackness at a rate of virtually 100%, but that something in the biological endowment of ravens makes it the case that they are black (and that the non-black ones represent mutations.) The same applies to weaker generalizations that do not invoke biological essences. Consider “college students drink beer.” College students as a kind do not only drink beer at a statistically high rate, but there is something in the way that college life is set up (social camaraderie/pressure, access to alcohol, etc.) that suggests the presence of a causal connection between collegiate status and beer consumption. Generics thus express something over and above what is expressed by the kind-wide quantifications that are their cousins. “100% of/All ravens are black,” and “80% of/most college students drink beer” certainly imply strong relationships between ravens and blackness and college students and beer consumption, but they stop short of implying that this correlation is nomic or non-accidental.

As it turns out, the idea that generics “imply” nomic or non-accidental correlations is not merely an armchair insight. A good deal of recent evidence from developmental psychology strongly suggests that the production and interpretation of generic language is causally responsible for both *triggering* essentialist beliefs as well as *socially transmitting* essentialist beliefs (Gelman 2003, Khemlani et al 2009, Rhodes et al 2012, Leslie 2014, Leslie 2017). We’ll go over some of these signature studies in some detail.

1.1 Generics and the social transmission of essentialist beliefs

First, exposure to generic language increases children’s tendency to endorse essentialist explanations for behavior. In one study (Cimpian and Markman 2009), children were presented with

a picture of a butterfly and asked to explain why it had a certain property. In one condition, children were asked to give an explanation for why a *particular* butterfly had the property, i.e. “*This* butterfly has dust on its wings. Why?” In the other condition, children were asked to give an explanation via a bare plural generic prompt, i.e. “*Butterflies* have dust on their wings. Why?” In the more particular condition, children were much more likely to explain the butterfly’s dustiness by making reference to possible extrinsic factors, e.g. “she flew through a dusty room.” In the generic condition, children were much more likely to appeal to system-based or functional explanations to explain the phenomenon, e.g. “they need dust to fly,” or “they need it for protection.”⁸⁹ Plausibly, the generic language condition led children to assume that what explains the presence of the trait among butterflies as a class is an underlying intrinsic entity or disposition common to all or most butterflies.

Of course, it is possible that these patterns of responses are better explained by features of children’s background knowledge about butterflies. In order to show that generic language itself is prompting these essentialist beliefs, we need to rule out the possibility that essentialist judgments and system-based explanations are products of this sort of background knowledge. Consequently, another series of experiments devised by Rhodes et al (2012) demonstrated that this link between generic language and essentialist beliefs also extends to novel kinds that children have not encountered previously. The experimenters made two picture books (with identical pictures) about a made-up kind, the Zarpies, whose members shared no obvious observable properties. In order to isolate the effect of generic language on essentialist beliefs, the authors controlled for other highly essentialized categories such as sex, race, and age:

To test whether hearing generic language induces social essentialism, we introduced

⁸⁹ Quantified prompts were excluded, since both children and adults tend to recall quantified statements as generics. See Leslie and Gelman 2012.

children and adults to a novel category of people—“Zarpies”—via an illustrated storybook, as in previous work on animal categories. Each page presented a picture of a single person displaying a unique physical or behavioral property. The characters were diverse with respect to sex, race, and age; thus, the novel category cut across groupings for which people might already have essentialist beliefs. For example, if all of the “Zarpies” were Asian, subjects might apply essentialist beliefs to the group because they generally have essentialist beliefs about race. Because the novel group is so diverse, it would initially appear arbitrary; thus, levels of essentialism in the absence of generic language should be low (as confirmed by comparison conditions). (Rhodes, Leslie, and Tworek 2012)

One picture book contained a high degree of particular language accompanying the pictures. For example, a picture of a Zarpie eating flowers would be accompanied by descriptions such as “This Zarpie eats flowers.” The second picture book contained a high degree of generic language. The same picture would be accompanied by a bare plural such as “Zarpies eat flowers,” or an indefinite like “a Zarpie eats flowers.” Children exposed to the generic text (read to them by their parents over a series of weeks) were more likely to give essentialist answers to questions like “This Zarpie was raised by a non-Zarpie mother. Will he prefer eating flowers or eating crackers?” Those exposed to the generic language tended to answer that the orphaned Zarpie would eat flowers, presumably since a Zarpie’s inherited essence would manifest itself despite non-Zarpie acculturation. Notice that this is a particularly strong test for essentialism, since children tend to think of flowers as bad to eat and crackers as good to eat.

This experiment strongly suggests that the production of generic language instills essentialist beliefs in interpreters. Yet there is also evidence that essentialist beliefs are causally responsible for the production of generic language (Rhodes et al 2012.) In a follow-up experiment, one group of parents was given a paragraph that purposely led them to form essentialist beliefs about Zarpies. The paragraph did not contain a high degree of generic language, but it did make claims about the “nature” of Zarpies, what they “essentially” are and do. The second group of parents was given a paragraph that focused less on what makes Zarpies a distinct species, and more on how Zarpies

behave. Then both groups were given the same picture book as in the original experiment, albeit with all text redacted. Parents were instructed to “talk through” the book with their children over a several week period. After this period, the parents and children both answered questions about Zarpies. As expected, the parents exposed to the essentialist language had a strong tendency to use generic language about Zarpies, as did their children.

This connection between generic language and essentialism may also help explain why “striking property generics” (Type B generics in Leslie’s taxonomy) are widely accepted even when there’s an extraordinarily weak statistical relationship between kind membership and the property in question. As we saw in Chapter 2, generics such as “Mosquitoes carry West Nile,” “Pitbulls maul children,” and “Sharks attack swimmers” are frequently accepted even in cases in which the “striking” property is rare among the kind under discussion. One hypothesis (Leslie 2007, Prasada et al 2009) is that these generics are accepted at low prevalence levels because the primitive generalization mechanism treats *overgeneralization* of a striking or harmful property among a kind as an acceptable tradeoff between cognitive effort and risk. After all, even if the chances of getting attacked by a shark are one in several million, from the point of view of self-preservation, it’s probably best to endorse beliefs that are likely to keep one out of the water.

However, it seems that something besides mere strikingness of the property accounts for the tendency to accept striking property generics at low prevalence levels. As Jennifer Saul (2018) points out, few people are tempted to accept generics like “office chairs explode,” although there have (bizarrely) been a handful of cases in which this has occurred. Similarly, few people are willing to endorse generalizations like “lawyers are rapists,” even though there is almost certainly a stronger statistical connection between being a lawyer and being a rapist than, say, being a mosquito and

carrying West Nile.

In response to Saul's point, one might suggest that interpreters tend to *overestimate* the presence of a striking property among a kind *unless* that inference is blocked by background knowledge concerning prevalence levels about the kind. Then the difference in acceptance between "office chairs explode" and "Muslims are terrorists" would be simply that everyone knows office chairs almost never explode, but that people are generally ignorant of the objective rates of terrorism among Muslims. Yet this clashes with the fact that interpreters *are* willing to accept striking property generics such as "Ticks carry Lyme disease," "Mosquitoes carry West Nile," and "Sharks attack swimmers," *despite knowing* that these properties are rare among each kind. For instance, Prasada et al (2009) asked adult participants to assess the prevalence of a given property among a kind in a wide variety of cases. Examples included estimating the prevalence of right-handedness among Canadians, the prevalence of Lyme disease carrying among ticks, etc. Subjects were fairly accurate in their estimates, guessing that about 60% of Canadians are right-handed (85% in reality), and that the vast majority of ticks do not carry Lyme disease. However, these prevalence estimates seemed to play no role in subjects' judgments about whether an *arbitrary* Canadian was right-handed, or whether an *arbitrary* tick carried Lyme disease. Subjects were just as likely to judge that *Joe the Canadian is right-handed* as they were to judge *Jumpy the tick carries Lyme disease*.

Background knowledge about a kind often does not block acceptance of certain striking property generics. There must be something else that explains why these generics are accepted. One hypothesis that we briefly discussed in Chapter 5 is that a striking property generic is particularly likely to be accepted when it concerns a kind which is *highly essentialized* (Leslie 2017). When certain kinds are believed to share in a group essence, kind membership is treated as particularly informative

about what sorts of properties the member possesses. When this group essence is believed to ground particularly striking properties, the connection between kind membership and striking properties is represented as non-accidental. This indicates that the sort of reasoning that underlies the acceptance and interpretation of Type B generics trades on essentialist beliefs.⁹⁰ Moreover, it's important to recall that this connection between genericity and essentialism appears to extend to inherently generic terms (recall the experiments in Gelman and Heyman 1999 and Waxman 2010, discussed in Chapter 4, section 5.1.)

To sum up, generic language has a number of connections to essentialist beliefs and explanatory strategies.⁹¹ First, experimental evidence suggests that generic language prompts interpreters to search for functional or system-based explanations for why an individual or kind instantiates a certain property. Generic language seems to function as a type of “attention-director” which prompts interpreters to focus on those properties which are perceived to be “intrinsic” or “inherent” to the kind. One consequence of this attention-direction is that generics decrease the salience of alternative “extrinsic” explanations of the phenomena under discussion. The obvious upshot for our current concerns is that generic language might play a vital role in raising to salience

⁹⁰ In Chapter 2, I argued that Leslie's semantics for generics did not have the resources to account for the meaning of Type C generics. I stand by that argument. However, these experiments strongly suggest that some Type C generics and Type B generics are very similar in terms of their psychological functions. Chapter 5 demonstrated that racialized categories are often essentialized in commonsense conception. Consequently, some of the Type C generics we have been discussing, including “Muslims are terrorists,” and “immigrants are rapists” might be accepted despite low prevalence levels because Muslims and immigrants are highly essentialized kinds.

⁹¹ Some recent work (Lazaridou-Chatzigoga et al 2017, Saul 2017, Hoicka et al 2018) has persuasively argued that generic language is not *special* in its capacity to trigger or transmit essentialist beliefs. The upshot of most of this work has been that putatively non-generic language, including kind-wide quantifications, is just as effective in transmitting essentialist beliefs. This work has been a valuable corrective to proposals in the literature that assume (Haslanger 2011, Leslie 2017) that if we were to eliminate problematic generic statements about social kinds, we would do much to eliminate problematic essentialist beliefs. Moreover, its thrust is compatible with my argument here: surely, racialized discourse can marginalize, oppress, and essentialize quite independently of its logical form. My point over the last few chapters is that *given* that much racialized discourse *does* come in generic form, we should expect those forms of racialized discourse to have the psycho-pragmatic effects characteristic of generics. That doesn't entail a commitment to thinking that “generic” racialized discourse is *special* in its capacity to essentialize.

essentialist explanations of racialized properties such as aggressiveness, criminality, and violence. Next, we will investigate the pragmatics of generic language, and show that generic racialized discourse has the capacity to do precisely that and more.

2. Generics and the Common Ground

In this section, we investigate the role that generic language plays in everyday linguistic exchanges. Philosophers have long conceived of linguistic communication as a rational, cooperative activity in which interlocutors' utterances contribute to a "common ground" of mutually shared information. Originally a term of Grice's, "common ground" has been most influentially and systematically explored by Robert Stalnaker's (1978, 2002) model of conversational context.

On Stalnaker's view, at any stage of a conversation, the participants in that conversation *presuppose* certain things. Presuppositions are propositions that conversational participants hold to be true. If *every* participant holds the same proposition to be true, that proposition is a shared presupposition. All shared presuppositions belong to a set called "the common ground." As the conversation develops, and interlocutors exchange information, the common ground-- the stock of information shared among interlocutors-- grows. The sort of information that speakers add to the common ground are *propositions*. Propositions are understood as sets of possible worlds. Roughly, to *assert* a proposition P constitutes a proposal to rule out the possible worlds which do not include P (that is, in which P is false.) If the assertion is accepted by other interlocutors, the worlds in which P is true remain in the common ground, and represent "live options" for further assertions.⁹²

⁹² The idea that interlocutors can *reject* or *resist* a proposed addition to the common ground is an important part of the Stalnakerian pragmatic picture. However, I largely drop discussion of this element in what follows. The psychological and pragmatic effects I claim for racialized discourse should thus be understood as going through on the assumption that no one rejects them.

For instance, in asserting “my dog is a Husky,” I propose to add all the worlds to the common ground in which 1) I own a dog and 2) that dog is a Husky. By the same token, I propose to rule out worlds in which I do not own a dog, or in which my dog is not a Husky, etc. These “subtracted worlds” are subtracted from what Stalnaker calls the *context set*. The context set consists of all the possible worlds in which all of the propositions included in the common ground are true. As common ground grows, the context set shrinks, since the addition of each new proposition rules out more and more worlds which were compatible with common ground prior to the update (we could imagine an idealized conversation in which *every* possible true proposition were stated and agreed upon by the interlocutors. In this case, there would be exactly *one* world in the context set, namely the actual world, and interlocutors would be unable to add any more true propositions.)

Stalnaker’s model provides an elegant account of how assertions can update conversational contexts. Since it is in many cases easy to grasp what sorts of worlds are compatible with assertions such as “my dog is a Husky” and which worlds are incompatible, precisely *which* worlds are being proposed as additions to the common ground and (therefore) subtractions from the context set is fairly unproblematic, both from the point of view of the conversation’s interlocutors, and from the point of view of the pragmatic theorist.

However, as we have seen in Chapter 2, assertions of *generic* sentences pose their own sorts of problems. Take the assertion “birds fly.” Plausibly, an interlocutor who asserts “birds fly” is not proposing that we add only those worlds to the common ground in which *all* birds fly. Why not? For one thing, “birds fly” can be true even if there are worlds (like ours) in which some birds *do not* fly. “Birds fly” thus can be a proposal to add at least some worlds in which fewer-than-all birds fly. Similarly, “birds fly” does not seem to be a proposal to add worlds in which *some* or even *most* birds

fly, since even a strong statistical correlation between being a bird and being able to fly would not be sufficient evidence for the claim that the assertion expresses: namely, that there is a lawlike or non-accidental connection between birdhood and flightedness. For these reasons, we might think that “birds fly” proposes something like adding all worlds in which birds *generally* fly, or *usually* fly, or in which *normal* birds fly, or in which *ceteris paribus* birds fly. And to be sure, one could defend any or all of these proposals, depending on what one believes about the semantics of generics-- for instance, perhaps normality theorists would want to characterize the assertion as a proposal to add all *normal* possible worlds, or as saying something about *normal* birds (Nickel 2016), whereas stereotype theorists may want to construe such an assertion as saying something about *stereotypical* birds (Geurts 1985).

However, any one of these responses seems to beg the question against its alternatives: it seems that any account of the pragmatics of generics which is not antecedently convinced by a particular semantics for generics (like the present account, which argues for neither normalcy nor stereotype semantics) is groundless. Perhaps, then, we should look for an approach which remains agnostic about these particular semantic issues but which also captures what sorts of information enters the common ground by means of generic language.

One proposal that seems to do exactly this draws on facts about how *non-assertional* speech acts such as imperatives seem to function within the common ground. A number of philosophers and linguists have pointed out that the pragmatic effect of imperatives like “shut the door!” can be modelled not as proposals to update the common ground with new assertional information, but as proposed *orderings on worlds* within the context set. “Shut the door” does not so much add a proposition to the common ground (although conversational participants will come to have the

belief that the speaker wants the door to be shut) as much as it carries the information that worlds in which the door is shut are preferable to worlds in which the door is not shut, at least from the speaker's perspective.

Drawing on this account, some (especially Veltman 1996) have proposed that generic assertions have a similar effect on the context set. To belabor the point once more, generics of the form *K's are F* are interesting assertions partially because their truth is compatible with the existence of individual *k's* which are not *F*. In other words, adding "birds fly" to the common ground is compatible with the existence of worlds in the common ground in which individual birds do not fly. But which worlds are added? And how does a speaker's grasp on these worlds explain how they can understand the generic sentence? One way around this puzzle is to treat generics not simply as proposals to add propositions to the common ground, but as proposed orderings on the worlds. Unlike imperatives, descriptive generics do not order the worlds in the context set according to a *preference* ranking. Rather, they order them according to an *alethic* ranking.

- 1) Birds fly.
- 2) College students like beer.
- 3) Anarchists are in favor of decentralization.

These sorts of generics propose that the worlds in which these kinds have a certain property are *more likely* than those in which they don't. This hypothesis is plausible in light of the fact that it nicely captures the sorts of things that interpreters actually infer from generics. Accepting "birds fly" leads interpreters to infer that flying is *normal* for birds, that an *arbitrary* bird is likely to fly, and perhaps (when certain conditions are met) that the capacity for flight is somehow rooted in the essence of birds as a kind.

Yet some generics seem to have different sorts of effects on the common ground. Namely, they seem to impose *preference* orderings without imposing alethic orderings on the worlds of the context set at all. Call this the class of *normative generics*.⁹³ Normative generics include:

- 4) Friends don't let friends drive drunk.
- 5) Boys don't cry.
- 6) A woman never leaves her man.

It is pretty clear that 4)-6) are false if they are interpreted as specifying what sorts of worlds are more *likely*. Given the evidence we have, worlds in which friends let one another drive drunk aren't intuitively more likely than ones in which they don't, and the same goes for crying boys and man-deserting women. In other words, worlds in which 4)-6) are true don't seem more likely than worlds in which 4)-6) are false. Of course, this is because there is a sense in which 4)-6) are not statements about how friends, boys, and women are, but statements about how they *ought to be*. On this recasting, normative generics don't impose *alethic* orderings at all. Rather, they impose *preference* orderings. Worlds in which boys don't cry or women don't leave their men are represented as preferable, or at least as closer to some ideal, than worlds in which boys cry and women desert their men.

⁹³ There is disagreement in the literature about how to best cash out the difference between “descriptive” and “normative” generics. One extreme option is that normative generics are not “true” generics at all, but merely statements of fact about rules and regulations (Cohen 1999). Closer to the mainstream, Leslie (2015) argues that the difference between normative and descriptive generics is that each type invokes different concepts. For instance, the lexical item “boys” might pick out *male children* in a descriptive generic like “Boys have X and Y chromosomes” and *ideal male children* in a normative generic like “boys don't cry.” Finally, Haslanger (2011, 2014) argues that descriptive and normative generics differ not with respect to their semantic profile, but with respect to their pragmatic effects. “Boys don't cry” contains a presupposition (or implicature) that refraining from crying is what boys *ought* to do, and “boys have X and Y chromosomes” contains a presupposition (or implicature) that there is a *statistical* relationship between being a male child and having a certain chromosomal structure. As we will see, Haslanger thinks that these effects are highly sensitive to context: interpreters might interpret “boys don't cry” as a statistical generalization, or “boys have X and Y chromosomes” as a normative one. However, for my purposes, nothing much hinges on what feature makes normative and descriptive generics different. The point of the moment is simply that generic language is capable of having two different effects on the common ground.

However, there is a third class of generics which seems to involve *both* sorts of orderings:

7) Shortstops are agile and speedy.

8) Students in this class write three papers.

9) Adult male Dobermans are between 85 and 100 pounds.

There is a clear sense in which these generics are “normative,” since they say something about what shortstops, students, and Dobermans *ought* to be like given the norms of baseball, philosophy classes, and canine health. Yet they are also “descriptive” in the sense that they say something about what shortstops, students, and Dobermans *are* like as a matter of fact. Importantly, these aren’t merely statistical relations between kinds and properties. Each generic says something about what is *likely to be the case* for each kind, where the relationship between kind membership and property instantiation is in some sense non-accidental.

Each of these generics expresses what is normal for a kind given some set of background norms. In contexts where these norms are presupposed, they manage to say something both about how shortstops, students, and Dobermans are, and also how they *ought* to be. In a moment, I will argue that racialized discourse can also have these common ground effects. But as we will see, we must take care in formulating this proposal.

3. Haslanger on Generics and Pragmatics

In recent papers, Sally Haslanger (2011, 2014) has proposed that a variety of generics lend support to harmful forms of ideology, including sexism, racism, and speciesism. Examples include “women are nurturing,” “Blacks are more criminal than Whites,” “women are submissive,” and “cows are food.” Her project is to understand how these generics support harmful ideologies by

virtue of their pragmatic effects. This project gives us an initial framework for understanding how the descriptive, normative, and normal dimensions of generic meaning interact.

Haslanger suggests that generics have three perspicuous effects on the common ground.⁹⁴ First, a generic *Fs are G* implies that *Fs are G* by virtue of something about what it is to be an *F*. This is especially true of comparative generics such as “women are more nurturing than men.” Here the suggestion is that it is something *about women as such* that makes them more nurturing than men (and not extrinsic or structural factors like different gendered socializations.) Second, generics often imply that these facts hold by virtue of the *essence* of the kind. As we saw in Section 1, there is a good deal of empirical evidence for both of these claims. Yet finally, Haslanger argues that generics can function as inference tickets from descriptive premises to normative conclusions. This is the claim we’ll examine for the rest of the section.

Haslanger begins by drawing attention to the familiar point that generic generalizations often seem to go beyond mere “statistical” generalizations. To belabor that point, “ravens are black” does not merely describe a statistically robust relationship between ravenhood and blackness, but rather describes what is in fact *normal* for ravens (Nickel 2008, Nickel 2016, Haslanger 2014). “Ravens are black” says that normal ravens are black (or that blackness is normal for ravens), and it says, or at least implies, that non-black ravens are abnormal. Consequently, knowing that *Fs are G* gives an interpreter a particularly rich source of information about *Fs*: if a given *F* is *normal*, it will be *G*.

Yet “normality” comes in at least two flavors. On one reading, “normal” is just what is *statistically* normal-- in conditions where there is a high or near-perfect correlation between being a

⁹⁴ Haslanger grants that these three effects are defeasible. We should only expect them to be in force within conversational contexts that are structured in a certain way, but Haslanger makes no attempt to specify necessary and sufficient conditions on what the conversational context must be like in order for these effects to manifest themselves (Haslanger 2014).

raven and being black, “Ravens are black” seems to capture something about what is statistically normal for ravens. Yet “normal” can also be what is *normatively* normal-- in conditions where there is an expectation that boys ought not to display emotion, “Boys don’t cry” seems to capture something about what is normatively appropriate or acceptable for boys. Of course, whether a generic is interpreted as saying something about statistical or normative normality is highly consequential. In general, it doesn’t make sense to hold individuals morally assessable simply for being *statistical* outliers. For instance, perhaps white ravens are *surprising* since they are statistically abnormal, but it requires further argument that their deviance from the statistical norm is grounds for punishment or correction. However, when individuals are *normative* outliers, they are deserving of sanctions. Despite their prevalence, boys who cry violate widespread norms concerning proper masculinity, and so are *defective*.

Haslanger’s hypothesis is that interpreters easily shift between these senses of “normality” when interpreting generic language. Moreover, the common denominator between statistical and normative normality is a conception of *essence*. As we saw, in many cases, a statistical correlation between being an F and being G can be explained by appeal to the essence of F.⁹⁵ Yet essence is also obviously at play in normative generics like “friends don’t let friends drive drunk”-- it is part of the essence or “nature” of a friend to not let their friends drive drunk. Drawing on evidence concerning the human tendency to essentialize natural and social categories, Haslanger argues that humans bring

⁹⁵ This is obviously not true of what we might call “purely statistical” generalizations like “cars have radios” or “barns are red.” There doesn’t seem to be anything in the essence of cars or barns that ground their observable properties. Or, to put it another way, the mere fact that these kinds and these properties are correlated is a sufficient explanation for why the generic strikes interpreters as being true. Of course, this simply may be because cars and barns, unlike ravens, boys, and women, are not highly essentialized kinds. See Prasada and Dillingham (2006) for more on generic taxonomy as well as data on how different sorts of generics are interpreted.

essentialist assumptions and *normative assumptions* to the common ground, unless they are contextually blocked:

Essentialist Assumption: Robust (meaningful?) regularities are not accidental. They are due to the natures of things.

Normative Assumption: Things should express their natures and under normal circumstances they will. Abnormal circumstances are not good and should be avoided or changed.

Now, the sorts of generics that Haslanger is interested in are ambiguous between statistical and normative interpretations. Generics like “women are more nurturing than men” or “women are submissive” certainly seem to capture certain statistical regularities, but they also seem to say something about how women are *by nature*. The implicature seems to be that women are submissive or nurturing *because* of their innate femininity. When circumstances are (normatively) normal, women will act in accordance with their nature-- and in general, that is a good thing:

Fs being G by nature isn't just how things are, it is how things should be. It is (normatively) normal. We who seek the good should help Fs be G, should regard Fs that are not G as defective and in need of correction... if correction is unsuccessful, then they are presumably defective and appropriately marginalized. (Haslanger 2014: 389)

One of the interesting things about Haslanger's proposal is that it leaves open the possibility that politically problematic generics are *true* in some sense. Generic language might *accurately* represent the world while simultaneously having undesirable cognitive effects. For instance, consider the things that speak both in favor and against asserting or believing a generic like “women are submissive.”

Similarly, under conditions of male dominance, women are, in fact, more submissive than men.

This is a true generalization and those who live under male dominance are justified in believing it. But again, if male dominance is hegemonic, this seems not only to describe how women happen to be, but more than this: how women are. (Haslanger 2011)

Haslanger's view thus doesn't recommend rejecting problematic generics on the grounds that they are *false*. Yet her worry here is that a generic like "women are submissive" says at least three things, depending on the context of use. First, it might say something about what is merely statistically normal for women. At this level, we can understand the relationship between womanhood and submissiveness as contingent or accidental, in the same way we understand the relationship between barns and redness. At first blush, there's nothing immediately politically problematic about believing that this statistical relationship obtains. Yet second, when the male-centric social world is taken for granted (as it frequently is), the generic also says something about what women are like at a *deeper* level-- it says that women are submissive as a function of the sorts of beings they are. This makes the connection between womanhood and submissiveness seem *nomiic* or *non-accidental*. Now, it's at least conceptually possible that this essentialist belief is neutral and benign. Yet the danger arises when this belief is linked to the *normative assumption*. Thus the third thing that the generic might say is that women are *properly* submissive. It is this reading which gives interpreters a *prima facie* reason to think that assertive women are abnormal or acting contrary to their essence. As we saw, Haslanger's suggestion here is that a "normative" conception of normality, when paired with certain hegemonic sexist assumptions, implies that "good" women are submissive, and that "bad" assertive women ought to be policed and corrected.⁹⁶

⁹⁶ It's worth noting that this is essentially Manne's (2017) conception of misogyny, which is understood as a punitive and corrective "police force" which targets women who are deemed out of line with the more "ideological" sorts of beliefs and impulses associated with sexism.

There is much to recommend Haslanger's story. First, her pragmatic hypothesis nicely ties together a number of things that are widely taken to be true of generics, particularly that they are linked to essentializing attitudes, that they express nomic or non-accidental generalizations, and that they can be ambiguous between "descriptive," "normal," and "normative" readings. Second, and more relevant to our current concerns, notice that we can leverage Haslanger's account to make sense of the idea that a generic sentence can impose an alethic ordering on worlds while *simultaneously* imposing a preference ordering on those worlds. When the essentialist assumption is part of the common ground, generics like "women are submissive" or "Asians are good at math" are interpreted as saying that women are submissive and Asians are good at math *because* of their essential nature. Presumably, because essential natures are widely cross-contextual and not easily lost, worlds in which women and Asians have the ascribed essence are intuitively more likely than those in which they do not. And when the normative assumption is part of the common ground, these generics are interpreted as saying that worlds in which women express their submissive essence and worlds in which Asians express their math aptitude are *preferable* to worlds in which they do not.

However, Haslanger's story can't be easily adapted to model the pragmatic effects of the sorts of racialized discourse we are interested in. The reason is simple: it is possible for a generic *F*s are *G* to trigger the essentialist assumption without triggering the normative assumption. We can bring this out by considering generics about things with salient bad or dangerous properties. "Natural disasters kill people" might trigger the essentialist belief that natural disasters kill people as a function of their intrinsic nature, but it's far less clear that a given disaster is a normatively *proper* natural disaster only if it kills people, and even less clear that natural disasters that don't kill people ought to be "corrected" so that they do. Equivalently, putting things in terms of the

common-ground orderings considered in the previous section, “natural disasters kill people” certainly seems to impose alethic orderings without imposing preferential orderings on possible worlds.

By the same token, Haslanger’s account as it is will not cover the types of racialized discourse we have been considering in this dissertation. Thugs, terrorists, immigrants, criminals, and welfare recipients may be essentialized, but they are also widely taken to be bad things with bad properties. Few think that it is a *good* thing that racialized groups express these natures. Consequently, we cannot assume that Haslanger’s normative assumption is part of the common ground when racialized groups and traits are under discussion. That assumption only seems to be in play when the putatively natural property is *good*, or at least neutral. The core limitation of Haslanger’s account, then, is that it cannot seem to make sense of the idea that *negative* properties can be regarded as both natural and preferable to alternatives. This suggests that we should investigate other explanations for how Type C generics and racialized terms introduce normativity into the common ground.

Of course, one approach would be to simply amend Haslanger’s account by tweaking the normative assumption to cover cases in which expressing an essential nature is negatively valenced or harmful-- that is, “bad” things should be *prevented* from expressing their natures. But there are good reasons not to do this. For one thing, the fact that many essential natures are regarded as having a negative or harmful orientation, and the fact that many of our generics concern harmful properties, jointly suggest that the normative assumption is highly defeasible. This puts pressure on Haslanger’s hypothesis that the normative assumption is a *default* feature of the common ground. Second, Haslanger’s hypothesis that the “normative assumption” is frequently taken for granted is already on shaky empirical grounds. Haslanger takes her hypothesis to be *compatible* with much of the

empirical evidence on the interpretation of generics, but that's very different from saying that it is *motivated* by that evidence. For instance, while Haslanger appeals to relevant evidence that suggests that "descriptive" generics are frequently interpreted as grounding normative inferences about how things ought to be (Prasada and Dillingham 2006, Prasada et al 2013), that evidence seems to underwrite a much more circumspect view about when such inferences are treated as legitimate.⁹⁷

Introducing a "negative normative assumption" into the common ground certainly is not licensed by this evidence. Of course, the value of Haslanger's work in this connection is independent of its empirical support. Yet I suggest that the inconclusive nature of that support enjoins us to seek a more independently motivated explanation of how generics introduce normative assumptions into the common ground. Beefing up the normative assumption in order to account for new data seems rather *ad hoc*.

4. Racialized Discourse as both Essentializing and Legitimizing

The central insight of Haslanger's approach-- that generic language can carry information both about what is the case and what ought to be the case-- is worth retaining. I still think that it's promising to think of generic language as licensing certain kinds of inference tickets from descriptive to normative conclusions (and vice versa.) Yet we've seen that Haslanger's particular account of

⁹⁷ In fact, much of this evidence (e.g. Prasada and Dillingham 2006) suggests that the normative assumption is much more likely to be in play when the kind is conceived as *obviously* natural. So, for instance, interpreters who accept "dogs have four legs" are likely to think that dogs have four legs because dogs simply are the sorts of creatures they are, and that dogs that do not have four legs are defective in some way. Yet generally, these two sorts of inferences are not drawn when unambiguously social kinds are under discussion: the truth of "police officers eat donuts" isn't generally taken to imply that police officers eat donuts because of their inner nature, nor that non-donut-eating officers are defective. Obviously, hybrid "natural/social" categories like racial and gendered categories are muddier, and running experiments on generics concerning these categories can be hampered by social desirability effects (which may even extend outside the lab.) But the fact that these categories aren't normally conceived as *obviously* natural-animal kinds suggests that we need a finer-grained study of when inferences to normative conclusions are treated as legitimate.

these inference tickets faces problems when it comes to the sorts of racialized discourse we are here interested in. In this final section, I will argue that (generic) racialized discourse is indeed capable of carrying these two types of information into the common ground, but in a more roundabout way than Haslanger suggests.

Recall the problem we are trying to solve. Assuming that the pragmatic function of generic language is to order the worlds of the context set along alethic and preferential dimensions, how can we explain these sorts of effects in the case of Type C generics and many racialized terms? In what sense does negatively valenced generic language like “thug,” “terrorist,” “Immigrants are lazy,” “Blacks are thugs,” “Muslims are terrorists,” and the like carries information about which sorts of possible worlds are both more likely *and* preferable to alternatives?

It strikes me as psychologically implausible that sentences such as “Blacks are criminal,” “Blacks are more criminal than Whites,” “Muslims are terrorists,” and unembedded racialized terms which carry racialized assumptions into the common ground are widely and explicitly interpreted as saying something about what sorts of worlds are preferable. One thing that seems clear is that even if interlocutors were to *prefer* that these states of the world obtain, interlocutors would likely not be *conscious* of that fact. They would likely reject that preference if it were made explicit. Another thing that seems clear is that if “Blacks are criminal” and “Muslims are terrorists” says something either explicitly or implicitly about what sort of world it is preferable to live in, it is curious that so much time and so many social resources should be devoted to *correcting* or *punishing* perceived Black and Muslim deviance.

I believe that solving this problem requires a change of tack. Haslanger is in effect trying to show that assumptions in the common ground play a role in determining the overall *content* of

generic propositions. When our assumptions are calibrated in the “right” (hegemonic) way, we can interpret “women are submissive” as saying something about how women ought to be, or how proper women are. Yet as we saw, this is a non-starter for many other instances of generic language. I thus suggest that we should not locate the “normative” element of racialized discourse in its fully fleshed out content, but in the *psychological function* of the sort of beliefs that racialized discourse transmits and triggers. We’ve already shown that by virtue of its semantic structure, racialized discourse introduces essentialist beliefs about racialized groups on the descriptive level. In closing, I suggest that racialized discourse imposes preference orderings on the common ground not by virtue of its content, but by virtue of the psychological role that those contents play. That is, *believing* that Blacks are criminal, Muslims are terrorists, Mexicans are rapists, etc. plays a role in justifying and maintaining the status quo, and it is the *status quo itself* that is taken to be preferable to alternatives. The concepts we’ll need to make sense of this claim are *status quo bias* and *system justification*.

4.1 *Status quo bias and system justification*

We can begin by distinguishing two senses in which A may be preferable to B. First, A might be preferable to B because A is a clearly good thing and B is a clearly bad thing. For instance, if you like ice cream but dislike asparagus, ice cream will clearly strike you as preferable to asparagus, all things being equal. Second, A might be preferable to B simply because A is preferable to any available alternative. Note that this does not require A to be a *good* thing. All that is required is that it seem better than B. For instance, imagine you are home alone watching TV. Watching TV is not entertaining to you; in fact, you are very bored by the activity. However, watching TV at home seems

preferable to reading a book, making lunch, or going to the gym-- those options are even *worse*, because they all sound exhausting. So, you continue being bored and watching TV.

The second notion of preferability is the one that is relevant to our purposes. In effect, when you stay home to watch TV, you are choosing to maintain the status quo in favor of alternatives. This is a manifestation of *status quo* bias. Those who manifest status quo bias need not think that the status quo is very good in order to maintain it; all that is required is that it seems *preferable to alternatives*. As it turns out, status quo bias is rather strong. The status quo often seems preferable to alternatives, simply because it is the status quo. For instance, evidence suggests that human interpreters will prefer events that are more likely *solely because they are more likely*. In one famous study, Kay et al (2002) found that leading up to the 2000 election, both Democrats and Republicans switched back and forth between preferring Gore and Bush presidencies, solely depending on how *likely* a Bush or Gore victory seemed. Jost et al (2004) describes this effect as a pre-emptive rationalization of the status quo “even before it became the status quo,” and generalize the hypothesis:

People will rationalize the (anticipated) status quo by judging likely events to be more desirable than unlikely events, a) even in the absence of personal responsibility, b) whether those events are initially defined as attractive or unattractive, and c) especially when motivational involvement is high rather than low. (Jost et al 2004)

Status quo bias is not merely an inertial or passive human tendency. Rather, it plays an active role in legitimizing those social systems. *System justification theory* is one approach to studying these legitimizing processes.

System justification theory begins with the idea that many of our beliefs are formed, maintained, and reproduced in response to specific psychological needs. Consider a simple example

of how a psychological need might affect belief. Imagine Adriane is cheating on Rachel, that Rachel knows this, and Adriane does not know that Rachel knows this. Each of them knows that if this fact becomes common knowledge, the relationship must end. Yet Rachel is deeply committed to the relationship, and can't bear for it to end. Consequently, over time, she manages to *convince* herself that Adriane is being faithful, even to the point of explicitly and ingenuously saying things like "Adriane would never cheat on me." In this case, Rachel's psychological investment in the relationship affects the sorts of beliefs she ends up forming in a rather direct way.

The relationship between psychological needs and epistemic states is not always this direct or dramatic. Yet it is rather commonplace. Psychologists and epistemologists often call processes like Rachel's mental gymnastics *motivated reasoning*. The idea is that reasoners are more incentivized to believe a proposition by their pre-existing commitments, investments, and desires than by what the evidence independent of their perspective may recommend. System justification theory focuses on how certain beliefs rationalize status quo social systems. Since status quo bias is pervasive, reasoners are often motivated to adopt patterns of belief which i) represent the status quo as better than it is, and ii) function to isolate social systems from reforming tendencies. Against this background, certain beliefs function not only as shorthand descriptions of social reality, but as *rationalizations* for social arrangements.

One potent source of system legitimization is a belief in "natural" social hierarchies. Even the most ardent defenders of the status quo must admit that the status quo is characterized by sustained patterns of inequality. In order to defend the status quo's legitimacy, these inequalities must be "naturalized." They must be seen as results of more fundamental sorts of inequalities between classes, genders, races, etc. rather than as engendered by status quo social practices and institutions.

This naturalizing impulse might lead believers to regard a variety of social inequalities as rooted in group essences rather than contingent social structures. In the case of racial inequalities, disparities in wealth, educational attainment, and incarceration among Blacks and Whites might be explained in terms of the differing “natures” of Blacks and Whites. To status quo defenders, the lack of motivation and inherent criminality of Blacks is what explains their overrepresentation among the poor and incarcerated, and the industriousness and law-abidingness of Whites is what explains their relative prosperity.

Evidence suggests that this naturalizing impulse is deeply connected to essentialist beliefs about race. Williams and Eberhardt (2008) found that thinking of race as biologically determined increased tolerance for extant patterns of racial inequality. Subjects who thought of race in primarily biological terms were more likely to think of racial inequality as “relatively unproblematic, natural, and unlikely to change via individual or social efforts” (Williams and Eberhardt 2008: 1036). In this case, biological thinking about race serves as a proxy for a racial essentialism which represents social inequalities as the natural expression of differences in group essence (Rhodes and Mandalaywala 2017). Similarly, Mandalaywala et al 2018 found that *both* Black and White adults who harbored essentialist beliefs about Blacks were i) more likely to be prejudiced against Blacks, and ii) were more likely to both endorse existing social hierarchies and to think that these hierarchies “reflect a naturally existing social structure.” The fact that this effect was observed across racial lines suggests that essentialism plays a role in system legitimization even when such legitimization seems to clash with group interest.

Finally, these naturalization strategies can exploit different conceptions of the group essence shared by members of racialized groups. Recall one study discussed in Chapter 5. Condit et al (2004)

found that interpreters high in racial prejudice were happy to explain disparities between Black and White IQ scores *either* in terms of “natural” or “cultural” causes, as long as these explanations focused on the traits of Blacks themselves. What seemed to matter to these interpreters was that this disparity could be rationalized in a way that put Blacks at the bottom of the social hierarchy, *regardless* of what mechanisms explained that position.

4.2 *The Psycho-Pragmatic Function of Generic Racialized Discourse*

If I am right that the online interpretation of racialized discourse implicates system-justifying patterns of belief, then we have a straightforward way of reconciling the hypothesized common ground effects of generics with a plausible psychological story of how interpreters come to believe that stereotype-consistent relations between racialized groups and racialized traits are both more likely and preferable. I have argued that (generic) racialized discourse *essentializes* its targets while playing a role in *legitimizing* the social hierarchies endemic to the status quo. These two functions correspond to the two pragmatic effects we have claimed are characteristic of generic language: *alethic* and *preferential* orderings of the worlds of the context set. If we pair these pragmatic and psychological effects, we can see that racialized discourse has an *essentializing-alethic* and a *legitimizing-preferential* psycho-pragmatic function.

Consider the first pairing. Type C generics like “Muslims are terrorists” and even unembedded racialized terms like “terrorist” can prompt interpreters to believe that a shared group essence non-accidentally links Muslim group membership to terrorist activity. Drawing on the evidence provided in Chapter 2, 4, 5, and section 1 of the current chapter, we can see that racialized generic language leads interpreters to focus on internal, essence-based explanations for racialized

traits. This has immediate upshots for understanding the pragmatic profile of racialized discourse. Generalizing the proposal, where the connection between a racialized group R and a property P is essentialized, “Rs are P” prompts the belief that Rs will be P in circumstances where normal conditions obtain. Worlds in which Rs are P by virtue of their essence will be intuitively regarded as more likely than worlds in which Rs are not P by virtue of their essence. This is the sense in which generic racialized discourse imposes *alethic* orderings on the worlds of the context set.

We should expect this essentializing-alethic function of generic racialized discourse to be linked to a variety of further inferences about the racialized kinds under discussion. As we saw in the discussion of “striking property generics” in section 1, we should expect an interpreter who accepts the generalization “Muslims are terrorists” to infer that an *arbitrary* Muslim is more likely to be a terrorist than an arbitrary non-Muslim, that *many* Muslims are terrorists, and perhaps even that *normal* Muslims are terrorists (i.e. that Muslims will realize their terrorism-prone essence unless their upbringing or environment interferes.) In some cases, we should expect these beliefs and further inferences to guide action and policy.

Now consider the *legitimizing-preferential* function of generic racialized discourse. While the essentializing-alethic function of racialized discourse says something about what racialized groups *are*, the legitimizing-preferential function says something about what racialized groups *ought* to be like, given a set of preferences. As we saw in our discussion of Haslanger’s understanding of the pragmatics of generics, any story about the connection between generic language and preferences needs to be told carefully. Haslanger’s hypothesis was that interlocutors bring a “normative assumption” to the common ground as a result of living under forms of hegemonic ideology. This explains why generics such as “Women are nurturing,” “Blacks are athletic,” and “Canadians love

hockey” can be interpreted both as saying something about how these groups *are* and how they *should* be. Worlds in which these groups have the properties attributed to them are not only more likely, but preferable to alternatives. Yet we saw that this proposal does not easily account for various forms of generic language that attribute *failures* or *pathology* to racialized groups.

To address this problem, I proposed that the relationship between generic language and preferences is less direct than Haslanger presupposes. Haslanger’s picture entails that preferences enter the common ground by targeting certain generic states of affairs-- *Women are submissive, Muslims are terrorists*-- as preferable to alternatives. This is what grounds the normative inference that (proper) women *ought* to be submissive. My picture, on the other hand, entails that preferences enter the common ground by targeting simply *what seems more likely*. Since one of the pragmatic effects of sentences like “Blacks are criminal,” “Muslims are terrorists,” “Immigrants are rapists,” and the like is that worlds in which these groups have the properties attributed to them are more likely than worlds in which they do not, these worlds are trivially preferable due to overwhelming status quo bias. If we understand status quo bias as a default feature of the common ground, we can see that a generic “Rs are P” does not produce an ordering according to which worlds in which Rs are P are *intrinsically* preferable to worlds in which Rs are *not* P (which is what Haslanger’s “normative assumption” commits us to.) Rather, Rs being P is preferable *simply* on the grounds that it appears more likely. Worlds in which Rs are P are thus regarded as preferable to a comparison class of worlds in which the relationship between Rs and P is otherwise than the status quo.

Finally, we saw that accepting and believing these generic contents is intimately bound up with patterns of *system legitimization*. Generic contents such as *Blacks are criminal, Muslims are terrorists, Immigrants are rapists, Poor people are lazy, Rich people are industrious*, etc. are powerful vehicles for the

transmission of system-justifying beliefs, since they encourage interpreters to believe that it is *something about Blacks, Muslims, immigrants, poor people, and rich people as such* that explains why they have their attributed properties. To the extent that social inequalities are seen as rooted in racialized group essences, generic contents provide quite direct legitimations of status quo social practices. More nuanced, system-based explanations for social pathology may effectively indict “the system” that plays a role in maintaining racialized inequalities. Yet since these explanations disrupt dearly held beliefs, worlds in which racialized populations can be represented as *responsible* for social pathology are defaultly preferred to alternatives.⁹⁸

5. Conclusion: Where We’ve Been, and What’s Next

The semantic profile of racialized discourse allows it to play a certain role in linguistic practices. Racialized discourse is characterized by race-related shifts in use, meaning, and conduct standards. Our semantic approach to racialized discourse so far has tried to explain how and why these shifts occur. Along the way, we’ve suggested that racialized discourse play a number of different roles in communication. Chapter 2 argued that the semantically relevant properties of racialized discourse have an irreducibly social dimension, and that it is this social dimension which explains why certain generic sentences display race-related meaning-shifts. Chapter 3 argued that racialized discourse has a powerful communicative function, both in the form of “code words” and in the form of “not-so-coded” racialized terms. Racial communication exploits the existence of race-related use, meaning, and conduct shifts in order to communicate substantive messages about racialized groups. In so doing, it re-entrenches these various shifts, whose mechanisms were

⁹⁸ There is obviously the possibility that the preferential-legitimizing effects of racialized generic discourse is a source of the moralistic attribution error discussed in Chapter 5.

characterized extensively in Chapter 4. Chapter 5 argued that various essentialist beliefs about racialized groups play a role in the online interpretation of racialized discourse, and that these beliefs give us a partial explanation of the psychological mechanisms underlying race-related shifts in conduct-standards. Finally, Chapter 6 argued that the semantic profile of some racialized discourse is particularly well-suited to transmitting and triggering the beliefs and essentialist explanatory strategies in Chapter 5. We've gradually built a case for thinking that racialized discourse has not only particular effects on what interpreters believe, but also affects what interpreters desire, value, and act upon.

5.1 The Argument So Far

All the way back in Chapter 1, we discussed Baltimore Councilman Carl Stokes' worry about using the term "thugs" to describe the young Black men rioting in Baltimore. Plausibly, Stokes' worry is that the commonplace use of terms like "thug" to describe these young men not only is a causal effect of an unjust racist history, but also that this sort of racialized discourse *itself* has detrimental cognitive and social effects. For instance, an interpreter who hears these rioters described as "thugs" may infer that Black youth are intrinsically violent or of irredeemably bad moral character. She might even infer that the social or political concerns of this group are not worth taking seriously, or even of excluding their concerns from public deliberation about "what is to be done" about a whole host of issues, from urban policing, to education, to local economies. In short, Stokes' worry is that in a wide variety of contexts, "thug" does some sort of socially marginalizing work.

In some ways, the conclusion of the argument of Chapters 1-6 is simply that Stokes is right to worry. Now we have a story that explains *why* he's right. Abstracting from the many details and qualifications of my account so far, we can sum these conclusions up simply as *The Argument So Far* (TASF):

The Argument So Far: The existence of racial domination opens up space for conversational contexts in which certain assumptions about racialized individuals and groups are widely taken to be true and normal. Racialized discourse transmits, triggers, and legitimizes these assumptions. As such, racialized discourse has effects on patterns of thinking, evaluating, and feeling about racialized groups which indirectly reproduce the conditions under which racial domination thrives.

On a more local level, TASF generates lots of more specific linguistic predictions of the following form:

TASF (for "thug"): The existence of anti-Black racial domination opens up space for conversational contexts in which certain assumptions about Black individuals and groups are widely taken to be true and normal, e.g. that Blacks are violent and aggressive. Racialized terms like "thug," and sentences such as "Tyrone is a thug," or "Blacks are thugs" can be used in ways that transmit, trigger, and legitimize these assumptions. These terms and sentences have effects on patterns of thinking, evaluating, and feeling about racialized groups which indirectly reproduce the conditions under which racial domination thrives through, for instance, material and symbolic support for aggressive judicial and executive policies that disproportionately disadvantage or oppress thuggish Blacks, unwillingness to live, work, or spend time in Black neighborhoods, approval for demagogic rhetoric about Blacks, etc.

We could run TASF in a similar fashion for a wide class of Type C generics ("Muslims are terrorists," "Immigrants are rapists"), as well as racialized terms besides "thug" (like "terrorist,"

“violent criminal,” and “welfare.”)⁹⁹ In the next, final chapter, we will do a little more to vindicate the idea that racialized discourse can reproduce the conditions under which racial domination thrives.

⁹⁹ Of course, I’ve only delved deeply into the semantics with regard to two sorts of racialized discourse, namely generics and i-level racialized predicates. So we might need a different account of the mechanisms that explain the truth of TASF for different sorts of vocabulary. However, while I do believe that the semantic details for those sorts of vocabulary would differ, I believe that my programmatic claim would still apply: racialized discourse brings about certain cognitive and social effects, and it does it by virtue of exploiting shifts in race-related use, meaning, and conduct standards.

Chapter 7. Constructing Racialized Categories

Abstract

In this final chapter, we turn to an investigation of how racialized discourse shapes the social world. My goal will be to offer a general account of how processes of racialization-- one species of which is racialized discourse-- socially constructs racialized categories. The notion of “social construction” I am presupposing here is what Ron Mallon (2016) has called “category construction.” According to category constructionists, social practices, beliefs, and widespread representations involving a category C “causally or constitutively explain how C came to have instances, or how C continues to have instances, or how members of a category came to have or continue to have their category-typical properties” (Mallon 2016: 147). In the first place, to say that racialized discourse socially constructs categories like *thug*, *terrorist*, *immigrant*, and *criminal*, is to say that racialized discourse (along with its attendant psychology) plays a role in explaining why these categories persist. Yet it is also to say something about how these particular categories gradually took on racial significances. We don’t just want to know why “thug” talk and thought makes it the case that there are *thugs*. We also want to know why “thug” talk makes it the case that so many members of the *thug* category are racialized as non-White.¹⁰⁰

While the analytic social constructionist literature has a good deal to say about how categories like *race* and *gender* are constructed by race-talk and gender-talk, it has had far less to say about extant social categories that gradually *become* racialized, and what the metaphysical significance

¹⁰⁰ This is not the same as saying that the sentence “Thugs are generally non-White” is always *true*. Rather, it is something like saying that membership in the *thug* category substantially co-occurs with membership in the *non-White* category *given the structure of extant social-racial practices*. As a number of thinkers engaged in “ameliorative” metaphysics have pointed out, accurate statements about what is the case within a particular social structure need not be understood as endorsements of those statements or that structure. (Recall Haslanger’s discussion of the truth of “women are submissive,” discussed in Chapter 6.)

of racialization might be. In order to address this lacuna, I propose that we should understand certain social categories as “shifty racialized categories” (SRCs.) SRCs are social categories whose membership conditions are defeasibly and contextually sensitive to the racialized properties of their members. Drawing on the historical case study of *criminal*, I show that racialized discourse can make it the case that a previously non-racialized social category ends up tracking racialized properties of individuals.

Words have no power to impress the mind without the exquisite horror of their reality.

--Edgar Allan Poe, *The Narrative of Arthur Gordon Pym*

0. Introduction

Much of this dissertation has focused on how individual minds and broader social structures manage to create and distribute *representations* of race and racialized traits. This focus on linguistic and mental representations has given us a good grasp on what is *said* and *thought* about racial kinds. Chapters 5 and 6 made the case that we should expect these sayings and thinkings to have effects on *actions*, and particularly actions which legitimize already-existing social hierarchies. What remains for us to show is that these sayings, thinkings, acting, and the broader social-historical context in which each of these occurs determines the metaphysical structure of the racial world.

The categories I have been discussing in this dissertation-- *thug*, *terrorist*, *criminal*, *welfare recipient*, and the like-- are what I will call “shifty” racialized categories (SRCs.) Unlike categories such as *Black woman* and *White man*, membership in an SRC is only defeasibly connected to one’s membership in a straightforwardly racial category. What makes the category a *shifty* racialized category, I contend, is that the conditions that an individual needs to meet in order to be a category

membership can shift from context to context, and often in a way that tracks membership in racial categories.

We thus need to explain two things about the relationship between racialized discourse and SRCs.

- *Category construction*: the theory should explain how racialized discourse constructs SRCs. That is, it should explain how racialized discourse makes it the case that certain social categories have *racialized* instances, where those racialized instances are non-accidentally linked to category membership.
- *Shifty grounds*: the theory should explain why which membership in a particular racialized category (*White, Black*, etc.) is only defeasibly connected to membership in an SRC.

In making sense of category construction, I suggest that bits of language can come to refer to specifically racialized populations when embedded within broader social and discursive practices. We first show this by way of historical example. Following Vesla Weaver's (2007) theory of "frontlash" in the formation of American crime policy, I show that the social category of *criminal* took on racialized significance as a result of an elite-led movement to establish control over "the Black problem" in the wake of Civil Rights gains. Next, abstracting a bit from Weaver's account, I provide a general model for understanding the metaphysical significance of racialization. I propose that category racialization has causal effects on what sorts of properties the category *tracks*. In particular, category racialization more or less covertly introduces new, race-sensitive conditions for membership in a particular social category. Understanding *criminal* as an SRC thus entails understanding it as a category embedded in social practices in a way that makes the racial features of individuals particularly relevant to whether or not they will be included in the category.

In making sense of the “shifty grounds” of SRCs, I propose that membership in many social categories (including SRCs) is *multiply realizable*. That is, there are many properties of an individual which, in any given context, are causally responsible for realizing that individual’s membership in a social category. This has a number of consequences. First, it means that an individual’s racialized properties *may* play this realizing role at some contexts, but not at others. Second, it means that many individuals who *lack* specifically racialized properties may satisfy other, “non-racialized” conditions for category membership. I then make use of Ásta Sveinsdottir’s “conferralist” account of socially constructed properties to make clear how certain properties and individual’s can “realize” category membership in particular contexts. I close by bringing this picture of the construction of social categories into alignment with my account of the semantics and pragmatics of racialized discourse.

1. Racializing Criminality: A Historical Example

To see how racialized discourse (broadly construed) constructs SRCs, let’s begin with a historical example of how the category *criminal* gradually took on racialized significances in the context of American politics. Here I draw on Vesla Weaver’s influential account of this process.

Weaver coins the term “frontlash” to describe political processes “by which losers in a conflict become the architects of a new program, manipulating the issue space and altering the dimension of the conflict in an effort to regain their command of an agenda” (Weaver 2007: 236). Frontlashes have three elements. First, they require “winners” and “losers” in a political conflict. When winners win an issue, they reform the status quo in ways that the losers resent. The losers then must find a way to secure their interests in this new status quo. Second, losers in a frontlash do not simply try to restore the status quo as it was before. Rather, they propose *new* issues and policies

which are conducive to their own interests while simultaneously attempting to *monopolize* that issue space. During this monopolization program, losers attempt to generate a particular common shared perspective or understanding of an issue while “discrediting competing understandings” of that issue (Weaver 2007: 236). Critical “focusing events” and the threat of crisis often shape which particular narrative losers try to peddle. Finally, when this monopolization program is successful, losers of the original conflict manage to “capture” the new issue.

Once “issue capture” has taken hold, it is highly disadvantageous to the original winners to go against this dominant understanding. Consequently, the original winners do not try to *revise* the understanding that the now-dominant issue entrepreneurs have molded, but rather position themselves *closer* to that dominant understanding. Further political maneuvers on the part of all parties now continue within the framework provided by the dominant understanding, and the frontlash is complete.¹⁰¹

Weaver uses the concept of frontlash to understand the growth of punitive racialized crime policy during and after the Civil Rights era. Let’s start with the basic conflict: the battle for Civil Rights for African-Americans. By the early 1960’s, anti-Civil Rights conservatives (including Southern Democrats) had made a strategy out of linking civil rights and crime. They often argued that civil rights demonstrations *constituted* criminal acts, or at least that such demonstrations *incited* criminal acts-- not least because of the propensity to violent crime common “among members of the

¹⁰¹ The notion of a “frontlash” is an obvious conceptual cousin to the notion of “backlash.” Weaver argues that while the notion of a racial/racist “backlash” has been the dominant way of accounting for reactionary racial anti-progressivism in the US, it offers a “pseudo-theory” which fails to specify i) the conditions under which we should expect backlashes to occur, ii) the aim or purposes of a backlash, iii) the role of elites and non-populist actors within backlashes, and iv) uses “backlash” as a shorthand for a diverse group of more fine-grained social forces such as “racial threat,” “symbolic racism,” or “moral panic.” A theory of frontlash is thus intended to supplant the dominant backlash concept, or at least to illustrate its limited utility in explaining changes in race-related political dynamics. (See Weaver 2007: 237.)

Negro race” (Weaver 2007: 241). Of course, this supposed propensity also underwrote conservative arguments in favor of segregation.

Conservatives of course lost the formal legislative battle over civil rights with the passage of the 1964 Civil Rights Act. Yet this did not defeat the underlying strategy to tie the problem of crime to race. Two “focusing events” rendered this strategy particularly acute. First, the national crime rate was objectively rising. Weaver canvasses a number of demographic trends and technical factors that might partially explain this increase in the crime rate. Yet it was the second focusing event that made *race* a central explanatory strategy in accounting for the increase in crime. The 1964 race riots in Harlem and Rochester, the Watts riot of August 1965, and the racial unrest that characterized life in major American cities long after the enactment of civil rights legislation provided a conspicuously racial “frame” for interpreting crime. Here, conservatives argued, was *evidence* that integration bred violent crime, as well as evidence that such lawlessness was a vital part of the movement for civil rights as a whole.

The confluence of a rising crime rate with high-profile “race riots” raised the profile of crime among policymakers at the federal level. While liberals and conservatives agreed that crime was a problem, they disagreed on what best explained it, and how to best address it. For liberals, crime was ultimately a problem of systemic “root causes” that generated material inequality and the conditions under which crime flourished. For conservatives, crime was ultimately a problem of insufficient enforcement of the law-- a “law and order” approach was needed to curb the crime wave. In the wake of the race riots, conservatives returned to their familiar strategy: racial struggle was not so much a *political* matter as much as it was a *criminal* matter, and those who were most responsible for disturbing “order” were African-Americans themselves. This constituted a direct

attack on the liberal “root causes” approach insofar as it located responsibility for the riots within Black political struggle and “street crime” itself, rather than the inequitable operation of social forces. As Weaver characterizes this strategy,

The frontlash relied on a two-stage maneuver. First, conservatives attached civil rights to lawlessness by arguing that civil disobedience flouted laws and would inevitably lead to more lawless behavior. Thus, nonviolent protest was connected to riots. But then, through a reverse claim, they disconnected the relationship they had just sewn, by arguing that the riots were not connected to legitimate grievances but to “crime in the streets....[S]upporters of punitive crime legislation now argued that civil rights strategies promulgated the idea that laws could be obeyed selectively. With the nation engulfed in violent protests, they reopened this argument to legitimize their attack on the civil rights agenda and initial appeals for heavier handed law enforcement. (Weaver 2007: 247)

One result of this slippery equation of crime with civil rights struggles was a comparative de-emphasizing of many of the most proximal *causes* of racial unrest-- mainly, police brutality against Blacks. This allowed conservatives to claim that fundamental “disrespect for law” was primarily driving urban patterns of violence and looting, and was perhaps even stoked by outside agitators such as Communists, hardened criminals, and “professional agitators.” In short order, conservative “law and order” ideology successfully captured the issue of crime. By “racializing crime” and “criminalizing and depoliticizing racial struggle,” conservatives were able to introduce race as a central component of good crime policy.

With the “root causes” approach to crime successfully discredited in the eyes of voters, liberals were faced with a dilemma. On the one hand, no liberal wanted to be seen as supportive of urban violence. Yet on the other, most liberals were supportive of civil rights gains, and wanted to continue to support civil rights movements. Thus, according to Weaver, liberals had two options. First, they could double-down on the “root causes” message, and say that urban riots were reasonable responses to “deplorable inner city conditions.” Second, they could agree with

conservatives that the riots themselves needed to be curbed, but disarticulate the rioting from the broader civil rights agenda. Yet in the wake of conservative issue capture, sticking to a radical root causes approach was not politically feasible. Thus liberals set about trying to separate the “real” civil rights movement from the “street crime” that found its expression in urban rioting.

This strategy failed for two reasons. First, the message that liberals espoused, which emphasized *both* law and order and racial uplift, proved unpopular, ambiguous, and open to merciless conservative critique. Second, since liberals had in effect *granted* that urban racial unrest did not have anything to do with “real” forms of civil rights protest, they effectively deprived themselves of a viable reason to *not* crack down on Black “lawlessness.” Liberals thus endorsed increasingly punitive measures. As the years wore on, the “civil rights” component of the liberal message was increasingly left to those with more radical political imaginations, and liberals began treating urban crime as a depoliticized technical management problem solvable via interventions such as gun control, family planning, and investment in education.

Yet by the 1970’s, the conservative orthodoxy on crime had become hegemonic, and this “technical problem” was addressed with increasingly sinister solutions. Once the link between race and crime had been forged via this frontlash, Black crime became a topic with significant political leverage. As we saw in Chapter 3, conservatives continue to raise the specter of non-White criminality in order to consolidate a base of racially resentful Whites. And as Weaver’s analysis shows us, once this dominant understanding of the link between race and crime took hold, conflicted liberals were incentivized to depoliticize and “manage” (Black) crime through interventions that telegraph a “tough on crime” approach. Thus as the post-Civil Rights Era gradually gave way to the War on Drugs and the current era of mass incarceration, it has become all-too-apparent that crime is

a racialized category, and that “racial” conduct is often “criminal” conduct. As Weaver’s analysis compellingly demonstrates, it is not surprising why modern liberals did so little to halt this process--once conservatives captured the issue of crime, the truly liberal alternative was simply no longer politically feasible.

2. Racialization, Category Membership, and Tracking

As of 2018, Blacks and Hispanics represent a combined 56% of the overall prison population, while representing a combined 28% of the total US population. Whites make up 33% of the total prison population while representing 64% of the US population. (Gramlich 2018) Blacks in particular are 6 times more likely than Whites to be incarcerated at some point during their lives. While Blacks have long been incarcerated at rates disproportionate to their representation in the general population, the gap between the number of incarcerated Whites and Blacks has steadily narrowed. In 1926, Whites made up 75% of the incarcerated population, and Blacks made up 23% (Langan 1991.) By 1986, the gap had narrowed to 53% and 46% Black, and today, Blacks outnumber Whites behind bars (Langan 1991, Gramlich 2018). Meanwhile, Blacks’ representation in the general population has remained fairly consistent, from 10% in 1926 to 13% in 2018.

We could of course appeal to a variety of social, economic, and political explanations for these disparities. These explanations (such as Weaver’s above) might offer detailed stories about the *causes* of racialization. Yet we now want to abstract from this sort of social-scientific approach in the interest of saying something general about the *metaphysical* import of racialization. Could the processes of racialization that the social scientist documents affect the nature and functional role of social categories such as *criminal*? Is there any reason to think that the pre-frontlash category *criminal*

on the one hand and the post-frontlash racialized category *criminal* are in a sense distinct social-ontological objects? As we will see, I think the answer to both of these questions is “yes.” I suggest that we can make sense of this claim if we understand how social categories manage to “track” different sorts of objects and properties over time.

Let us begin with an obvious point: despite the widespread notion that punishment is a legally sanctioned *response* to criminal activity, legal and executive institutions often do not respond *only* to objective patterns of law-breaking. Consider the fact that police officers often “look the other way” when they observe minor infractions, or the fact that the same infraction (e.g. illegal marijuana use) can garner different responses depending on whether the lawbreaker is a White college student in a well-to-do neighborhood or a Black teenager in a working class neighborhood. This judicial distance between crime and punishment even holds for more serious sorts of crimes. One 1991 study conducted in the midst of the American carceral boom (Thomas and Boehlefeld 1991, quoted in Davis 1998) found that an extremely small subset of all “victimization offenses” resulted in a perpetrator’s contact with the criminal justice system.

Although annual statistics may vary slightly, a general trend over the past decade is clear. Less than 40% of victimization offenses are reported to police, and only about 20% of known crimes are cleared by arrest. The CJS Sourcebook shows that of those arrested, about 80% are prosecuted, three-quarters of those prosecuted are convicted, and about 70% of all felony convictions result in a prison or jail sentence. Consequently, only three persons are incarcerated (in prisons or jails) for every 100 crimes committed. (Thomas and Boehlefeld 1991: 246)

These considerations suggest that it’s not quite accurate to think of punitive apparatuses as *detectors* of “criminality.” The obvious and well-documented selectivity and bias within the criminal justice system stretches that metaphor. Rather, we should rather think of these apparatuses as performing certain *management functions* on whatever groups seem to be “problem populations” within a given

sociopolitical moment. The profile of these problem populations is highly variable. Prior to the punitive frontlash, “before 1960, crime was about the drunk, the ‘mad-dog sex killer,’ and the juvenile hoodlum... When crime was discussed, there was no explicit and very little implicit connection of crime to blacks.” (Weaver 2007: 242) Punitive apparatuses thus require “filters” in order to narrow down which populations are problematic and which are not.

As Angela Davis points out, race is often a “filter” for determining just *what* sorts of conduct will be criminalized, and just *which* sorts of perpetrators will be subject to legal sanctions. Davis notes racialization of crime often serves the purpose of re-securing White control over Black labor when social and economic forces render that control unstable. For instance, in the aftermath of the American Civil War, most American Blacks entered a nominally “free” labor force for the first time. Yet through a legal reinterpretation and expansion of what sorts of conduct constituted criminal offenses, free Blacks were often subject to different legal standards than their White counterparts:

The swift racial transformation of imprisoned southern populations was largely due to the passage of Black Codes, which criminalized such behavior as vagrancy, breach of job contracts, absence from work, the possession of firearms, insulting gestures or acts. The [1865 and 1866-- *PO*] Mississippi Black Codes, for example, defined a vagrant as “anyone who was guilty of theft, who had run away [from a job, apparently -*AD*], was drunk, was wanton in conduct or speech, had neglected job or family, handled money carelessly, and... all other idle and disorderly persons.” In other words, white behavior that was commended and thus went unnoticed by the criminal justice system could lead to the conviction of black individuals and to the ideological criminalization of black communities. (Davis 1998: 100)

As Davis and many others since have since shown (e.g. Blackmon 2008), this selective criminalization of Black conduct served a variety of economic ends, not least of which was the expansion of the convict-leasing system. Under this system, incarcerated Blacks effectively became an unpaid labor force, and perhaps an even less protected class than enslaved Blacks.

My suggestion is that these racializing processes affect just what bits of the environment social categories actually end up tracking. The examples we have been working with-- *criminal, terrorist, immigrant, thug, welfare recipient*, and the like-- show that a social category might track a number of properties, only some of which are racialized. Determining *when* these racialized properties are playing a causal role poses a number of normative and epistemic problems. My goal here is simply to get clear on the conditions under which these properties *could* play a causal role, and how those properties might figure into informative, true explanations of why the social-racial world is constructed the way that it is. We thus begin with a characterization of what it would mean for a social category to “track” a property.

2.1 *The Tracking Relation*

Let us say that a social category C “tracks” a property P under the following conditions:

Tracking: Category C tracks a property P iff the fact of having P non-accidentally co-occurs with membership in C.

Much more would have to be said in order to make this definition precise. Here I will simply specify in a bit more detail what it would mean for the possession of a property P to “non-accidentally co-occur” with membership in a category C. The first thing that the idea of “non-accidental co-occurrence” captures is the notion of supporting certain counterfactuals. If property P non-accidentally co-occurs with membership in category C, *not* having property P increases the likelihood of *not* being in category C. Obviously just which counterfactuals are supported will differ on the properties and categories involved. For instance, lacking the property of being *unmarried* is sufficient for not belonging to the category *bachelor*, but lacking the property of being *at least 60 years*

old merely decreases the probability of belonging to the category *retiree*.

The second thing that non-accidental co-occurrence captures is the idea that there exists some *mechanism* which makes it the case that a certain property tends to co-occur with membership in a given category. This is already a familiar point from our discussion of generics in Chapter 2. “Ravens are black” strikes us as true not only because ravens are in the main black, but because blackness is *normal* for ravens (Nickel 2016). The correlation between ravenhood and blackness is non-accidental to the extent that natural and biological mechanisms tend to produce black ravens. Of course, the mechanisms for these correlations need not be so robust, and they need not be natural or biological. In some cases we might be interested in weaker sorts of *social* explanations for what seem to be non-accidental correlations between *Greek immigrants* and *diner ownership*, or *Kenyan national origin* and *marathon success*, or *Anglo-Saxon descent* and *intergenerational wealth*. As these latter examples suggest, the mechanisms involved in an adequate explanation of these co-occurrences might be complex and richly relational. Yet what is important is that “non-accidental co-occurrence” tends to denote a rather *robust* relationship between a property and category membership.

Drawing attention to these two features of non-accidental correlations is already enough to get some grasp on what it would mean for a category *not* to track a property. First, it indicates that a category will not track a property where that property is in-principle *irrelevant* to an explanation of their membership in C. For instance, imagine that every instantiation of the *hipster* category at a given time or place happened to have a driver’s license ending in an odd number. Given that there is likely no non-accidental relationship between odd-numbered licenses and being a hipster, we can rule out that the *hipster* category tracks odd-numbered licenses. Second, it indicates that a category will not track a property where the relationship between that property and category membership is not

suitably *robust*. For instance, there will sometimes be one-off or even anomalous relationships between a property and category membership. For instance, imagine that Keiko is a high-school student who, due to clerical error, is admitted to Harvard, even though she never applied. Even though the category *Harvard undergraduate admittee* certainly tracks the property *high school student*, the mere fact that Keiko is a high school student does not explain why she is a Harvard admittee. In this case, the mechanism that helps to make it that case that Harvard admissions track high school status-- namely, the application process-- is subverted. This sort of case thus should not be understood as an instance which speaks in favor of a tracking relation between Harvard admissions and high school status.

More positively, note that a category may track *several* properties. Which tracked properties “stand out” to us will depend on what sorts of mechanisms we are trying to uncover. Similarly, the tracking relation is compatible with there being different “strengths” of co-occurrence, depending on what sorts of co-occurrences we are treating as significant. While this does mean that we can be relatively pluralistic about what sorts of properties a category tracks and the nature of their co-occurrence, it does not mean that we can be relativists about the tracking relation itself. That is, I take there to be facts of the matter about whether a category tracks a property, regardless of whether or not anyone has uncovered the particular mechanisms that explain the co-occurrence of the property and category membership. For instance, authorities and institutions may sincerely hold that their *terrorist* ascriptions track objective patterns of lawless intimidating violence around the globe. That is, it is these objective patterns which figure into the best explanation of why the category *terrorist* has the instantiations it has. However, it is plausible that an explanation of why the *terrorist* category generally has *these* instances (e.g. mostly *Muslim* and/or *Arab* instances) and generally does

not have *those* instances (e.g. very few *Christian* and/or *White* instances) must entertain the possibility that *terrorist* ascriptions (also) track racialized properties.

This consideration suggests that there may be multiple properties that a category tracks, and that some of these tracking relations may be more or less “officially endorsed.” Recall Haslanger’s distinction between manifest concepts and operative concepts, discussed in Chapter 2. The manifest, officially sanctioned version of the category *terrorist* might track all and only those individuals who engage in unilateral forms of vigilante violence and intimidation for political and ideological ends. However, the operative, implicitly practiced version-- the way the category “actually functions” within a set of social practices, perhaps independently of its manifest characterization-- might track all and only those *Muslim* individuals who engage in “terroristic” forms of violence.

Throughout, I have been making certain assumptions about what sorts of properties racialized categories do and do not track. For instance, I assume, without argument, that the category *criminal* does not track the property *White*, but does track the property *Black*. This might seem odd, since the minimal characterization of the tracking relation given above certainly seems to leave open the possibility of there being a non-accidental correlation between being White and being subject to criminal penalties, incarcerated, and so on. Yet I believe that my focus on certain tracking relations, as well as my assumption that certain tracking relations do not obtain, is justified on methodological and political grounds.

First, understanding how the social world is constituted under the conditions of racial domination may require us to entertain the possibility that members of systematically privileged groups are *not* non-accidentally linked to membership in marginalized social categories. I here assume that racism is an interlocking system of norms and practices which produces systematically

(and unjustly) differential life chances for individuals and groups based on the racial presentation of those individuals and groups. What makes those chances *systematically* differential is the existence of mechanisms which ensure that differential race-specific outcomes are “more than accidental.” Thus to say that a category like *criminal* tracks the property *Black* but not the property *White* is effectively to say that from the perspective of the norms endemic to a racist system, Black criminality is a “feature” while White criminality is a “bug.”

Second, and relatedly, mapping the social-racial world under the conditions of racial domination might require us to narrow our focus to specific co-occurrences which are the likely products of such domination. For instance, on Paul Taylor’s (2004) approach to racial ontology, understanding what the category *race* fundamentally is requires us to investigate the specific ways in which an individual’s racial presentation is linked to facts about that individual’s social position and life chances. In many cases, this requires us to focus not just on *any* correlations between race and social positioning, but on those non-accidental correlations that seem to flow from racial domination and injustice. As Taylor puts it:

A person we’d call black... is more likely to live in substandard or overcrowded housing, or lack health insurance, or be unemployed, than someone we’d call white.... This same person is more likely to be, or to have been, in the criminal justice system or to perform less well in school. And this makes some people worry, since it seems to invite us to draw conclusions about black inferiority. But the falsity of classical racialism prevents us from taking these correlations as evidence of some congenital incapacities or tendencies and requires that we look instead for social explanations...*because they’re black* is no longer an explanation for anything. It becomes, instead, a gesture at a request for an explanation, or for an answer to a question like this: What is it that links black people to these social conditions? (Taylor 2004: 85)

Taylor’s view is that *racess* are social roles of some kind. It is the existence of these roles which explain why these co-occurrences between bodily appearance, ancestry, and life chances persist. I believe

that we should be ontologically committed to SRCs for analogous reasons: it is the existence of the category *criminal*, or *terrorist*, along with the way such categories are embedded in our social practices, which explains why their membership conditions so systematically overlap with racial membership conditions.

It's worth noting that the view Taylor and I share is more broadly represented in the analytic social constructionist literature. For instance, Ron Mallon argues that we need to posit "social roles" in order to explain the "co-occurrence" of certain types of facts with other facts quite generally:

We do not want merely a common explanatory component of the individual properties [of a system], but of their reliable co-occurrence and co-instantiation-- of why the skunk-typical properties co-occur in skunks... we neither want an explanation of the biological traits nor of the (ex hypothesi) constructed traits, but of their co-occurrence in particular places and time. According to the social constructionist, the social role explains this co-occurrence. (Mallon 2016: 201)

2.2 *Multiple Realizability*

Recall the features of SRCs we are trying to explain. First, we want to explain how it is possible for there to be a non-accidental relationship between membership in a social category and the possession of certain race-specific properties. Second, we want to explain how it is possible for the race-specific properties of an individual to be non-accidentally linked with membership in certain categories without those properties *fully determining* which properties the individual belongs to.

We have already given a provisional explanation of the first feature of SRCs. A category C can be classified as an SRC when there is a shift in the nature of the properties that C "tracks"-- in particular, C begins to track specifically racial properties. As we saw in Weaver's and Davis' analyses of the racialization of criminality, the particular processes by which these tracking relations are forged are causal and historical in nature. Yet we can also understand racialization as having a

metaphysical role. In particular, racialization processes and the practices in which those processes are embedded provide the conditions for a non-accidental co-occurrence between membership in a social category and one or more racialized properties.

In this section, we take on the second feature of SRCs. Here we need to demonstrate the compatibility of two claims: 1) that a category C tracks a racialized property R, and 2) that R is not necessarily causally responsible for an individual I's membership in C. That is, the connection between the racialized properties of an individual and membership in an SRC that tracks that property is defeasible. At this point, we should distinguish between two questions we might ask about social categories. First, we might ask what sorts of properties a social category C tracks. To ask this question is to investigate (usually *a posteriori*) which properties tend to co-occur with membership in C, and to investigate whether there is any interesting mechanism which explains this co-occurrence. Second, we might ask what sorts of properties instantiated by an arbitrary *individual* explain their membership in C.

It is important to recognize how distinct these questions are. Consider an example. As public health professionals, we might be interested in what sorts of properties and activities tend to cause *lung cancer*. One thing that especially jumps out is that lung cancer seems to “track” habitual smokers (tobacco use accounts for about 90% of lung cancers.) This robust, group-level co-occurrence between membership in the category *lung cancer patient* and the property of being a smoker suggests that some mechanism is responsible for this correlation. On the other hand, we might be interested in what makes it the case that an *individual* patient, Cal, has lung cancer. When we investigate the details of Cal's case, we may find that he doesn't instantiate the tracked property we expected to find-- while he has lung cancer, Cal has never smoked a cigarette. The properties that are causally

responsible for Cal's condition may be any *other* number of lung-cancer-tracked properties, or even *none* of the lung-cancer-tracked properties we are familiar with. Cal might be a medical anomaly: he might have developed lung cancer by having a certain genetic endowment that we did not even know was tracked by lung cancer.¹⁰²

Similarly, we might ask what sorts of properties the category *incarcerated felon* tracks. Demographic data suggests that incarcerated felons tend to be poor, undereducated, and disproportionately non-White, and there are good reasons to think that there are enumerable social, political, and economic mechanisms that explain these co-occurrences. However, it is certainly not the case that *every* incarcerated felon instantiates all or even any of these tracked properties. After all, the White, the well-educated, and the wealthy occasionally find themselves behind bars as well.

What this suggests is that there are *multiple*, possibly disjunct properties, the possession of which could be causally sufficient to rule an individual into a social category. In other words, category membership is *multiply realizable*. Membership in a social category C is multiply realizable to the extent that there are many distinct properties pertaining to an individual I which could be causally responsible for I's belonging to C.

Note that the properties that realize an individual's membership in a social category need not be properties that the category *tracks*. Recall the case of Keiko the accidental Harvard undergraduate. The properties that realize Keiko's membership in the category *Harvard undergraduate admittee* are

¹⁰² Note that according to my characterization of tracking above, lung cancer would track Cal's genetics just in case there was a non-accidental, mechanistic link between his genetics and lung cancer. This risks trivializing the robust relationship I take to hold between racialized properties and membership in certain categories, since it seems to entail that as long as there is a story about *why* an individual belongs to a given category, the category tracks whatever property answers that question. Yet this depends on the details of the case. As mentioned above, the relationship between Cal's genetics and lung cancer may not be very *robust*. That is, perhaps 100 identically genotyped twin-Cals, raised in different environments, do not manifest cancer. Category C tracks property P" thus does not obviously collapse into "Category C tracks whatever property P of individual I is sufficient to include I in C."

something like *high school student* and the property of being *mistakenly believed to have applied to Harvard*. As we saw above, while *Harvard undergraduate admittee* tracks *high school student*, here the relationship between Keiko's high status and her admission to Harvard is "accidental" in a way that makes this tracking relation somewhat irrelevant to the present case. And there isn't any reason to think that *Harvard undergraduate admittee* tracks anything like the latter property (unless the admissions office starts malfunctioning in an impressively systematic fashion.) Nevertheless, it is some combination of these properties that are in fact causally responsible for Keiko's admission.

We now have in hand the notion of a category-specific *tracked property* and the notion of a category-specific *realizer property*, and can see that they are conceptually distinct. Call a realizer property a *base property* just in case it is a property that is in fact causally responsible for realizing an individual's membership in a social category at a context. As before, the base property that realizes an individual's membership in a social category may or may not be tracked by that category. This possible divergence between tracked properties and base properties gives us a straightforward way of explaining why social categories often have *atypical* or *anomalous* members. Keiko may have gotten into Harvard, and Jim the White stockbroker may have been incarcerated for financial crimes, but since the base properties that realize their respective statuses are atypical for the category, we need not think of this as being evidence for a tracking relation.

2.3 Multiple realizability and social construction

Thinkers sympathetic to social constructionism have already appreciated the utility of the multiple realizability thesis in making sense of the membership conditions for social kinds such as *race* and *gender*. One of the major motivations for the multiple realizability thesis within the

philosophy of mind was that it offered a way of making sense of the idea that mental states (e.g. pain) are causally realized by physical states (e.g. neurochemical states) without being fully reducible to those states. This anti-reductionist strategy is also useful for social constructionists attempting to map the different ways in which race and gender categories can be constructed across social contexts.

Consider the example of *race*. According to social constructionists, racial categories are not individuated by any biological or essential features, but by social and historical features. That is, races are somehow products of our social practices, particularly by the way in which we have decided to assign social significances to diverse body types. “Racial” classifications thus rely on a loose set of criteria having to do with physical appearance, ancestry, and other factors such as self-identification and cultural affinity (Mills 1998). Yet since the nature and scope of these race-producing social practices often differ between geographical and historical contexts, and because (ex hypothesi) these differing practices are authoritative over what racial groupings exist at their respective local contexts, this entails that both the racial groupings themselves and the criteria for classifying individuals into these groupings will differ. So, for instance, light-skinned Blacks in the United States are likely to count as “White” in present day Brazil, and an Irish-American immigrant arriving on US shores in 1840 would count as a member of a not-quite-White “mongrel” race, but as unmistakably “White” in 2019. In sum, as a well-known social constructionist slogan has it, “race does not travel.”

Similarly, many feminist thinkers and activists, often motivated by an inclusionary politics for queer and trans* individuals, have argued that membership in the category *woman* is determined neither by the actual possession of biological features such as chromosomes and genitalia, nor simply by being widely *believed* to possess these features (e.g. Jenkins 2016, responding to Haslanger

2000). It may be that the membership conditions for *woman* are highly context-sensitive, and that different properties of individuals may be relevant to category membership at different contexts. For instance, at some contexts (such as everyday social interactions with friends and peers), the fact that an individual *self-identifies* as a woman will be sufficient for her to belong to the category. At other contexts, however (such as stating one's gender on an application for a passport), self-identification as a woman will likely be insufficient to satisfy the operative standards for category membership in that context.¹⁰³

Sally Haslanger (2000) relies on the multiple realizability thesis when she points out that membership in a *gender* category is best thought of not as a matter of actually having certain intrinsic features, but as being *viewed* as having certain features mapping onto accepted sex-coded social roles. Since the nature of these features, the nature of the roles involved, and their assumed significance can vary between contexts, Haslanger suggests that a descriptively and politically adequate account of *gender* will recognize that membership in gender categories is “realized in different ways in different contexts.” Such a metaphysics permits us to make informative meta-level generalizations about how gender systematically affects the social standing of its category members without making reductive assumptions about the properties that “all men” or “all women” have:

Critiques of universalizing feminisms have taught us to be attentive to the variety of forms gender takes and the concrete social positions females occupy. However it is compatible with these commitments to treat the category of gender as a genus that is realized in different ways in different contexts; doing so enables us to recognize significant patterns in the ways that gender is instituted and embodied. (Haslanger 2000: 38)

¹⁰³ Again, this isn't to say that the passport office's standards for *woman* are correct or defensible. Correctly pointing out that self-identification as a woman is insufficient to meet those standards might be an example of getting a “good” (correct) read on a “bad” (unjust) social ontology. Here the goal is description of the ontology, not necessarily amelioration (see footnote 100.)

Similarly, Aaron Griffith (2018) points out that treating race and gender as multiply realizable kinds gives us a straightforward way of accommodating the social constructionist commitment to thinking of race and gender as relatively plastic sorts of social kinds not defined by any intrinsic feature. The question of whether an individual belongs to a given race or gender category is often not reducible to the question of whether that individual possesses any *particular* property that realizes category membership--for instance, being *Black* or being a *woman* is not always a matter of having a certain skin color or certain genitalia. However, if category membership is multiply realizable, particular individuals who count as *Black* or as *women* may possess *other* properties which, within a particular social context, are sufficient to realize their membership in these categories. The multiple realizability thesis thus offers a non-reductive account of category membership, as well as an explanation for the diversity *within* a specific social kind:

The multiple realizability of social kinds guards against their reduction to particular social features. Because being a K can be realized by different social features of different individuals, the kind K is not identical to any particular realizers. Moreover, the multiple realizability of social kinds also affords us a way of accommodating differences between members of the same social kind.

(Griffith 2018: 250)

My appropriation of the multiple realizability thesis to make sense of the membership conditions of SRCs is not novel. Yet it does offer a promising way of articulating the precise sense in which these categories are *racialized* without it being the case that category membership is *defined in terms of* having particular racialized features.

2.4 *What SRCs are*

We now have in hand the materials we need for a more precise characterization of “shifty

racialized categories.” First, SRCs are categories which *track racialized properties*. Second, membership within an SRC is *multiply realizable*. Finally, racialized properties are especially *salient* realizers for category membership.

Thus to say that a category such as *criminal*, *terrorist*, or *immigrant* is an SRC is to say that there is a significant and non-accidental co-occurrence between racialized properties and membership in the category. Racialized features of *individuals* can then be seen as playing a causal role in realizing their membership in these SRCs. Racialized properties are possible *realizers* for category memberships. Since category membership is multiply realizable, there are many ways to satisfy the membership conditions associated with *criminal*, *terrorist*, *immigrant*, etc. A given set of racialized properties thus may or may not be a part of the causal story of why an individual satisfies those conditions at a context. Moreover, there will often not be any category-specific property-- racialized or otherwise-- which realizes an individual's category membership in *all* contexts.

The multiple realizability thesis allows us to accommodate a number of features of SRCs. First, since there are many routes to category membership, the thesis explains how a category might track a racialized property without it being the case that *every* category member has that property. Second, it explains why the question of what properties are actually causally responsible for the fact that an individual belongs to a given category membership can be *overdetermined*. For instance, there may be individual Blacks who are *criminals*, but who also instantiate other properties which would also be causally sufficient to include them in the category *criminal*.

Finally, note that whether a racialized property *will* in fact realize category membership at a context (that is, whether or not race-specific properties will turn out to be the *base properties* causally responsible for including an individual into the category) substantially hinges on how things stand

with the surrounding social environment. Processes of racialization such as the “Islamization” of terrorism or the progressive color-coding of criminality provide the social conditions under which these categories can track racialized base properties. This dissertation has focused on some of the concrete linguistic and psychological mechanisms by which the social world takes on racial significances, but these should be understood as just one part of a much more complex story. One methodological upshot to the sort of social metaphysics I am pursuing here is that it must be substantially informed by historical and empirical work on whether and how social categories acquire racialized significances.

3. Conferralism and Racialization

In this final section, we have two goals. First, we want to capture in more detail the specific mechanisms by which an individual’s racialized properties might play a role in realizing their membership in certain social categories, particularly SRCs. Second, we want to understand why racialization is *metaphysically* significant as opposed to merely *epistemologically* significant. That is, we want to understand how racialization can play a role in determining what SRCs and other racialized categories actually *are*, rather than “merely” playing a role in determining how SRCs and other racialized categories are *represented* by individuals.

In meeting the first goal, I will claim that the properties that realize membership in SRCs are assigned from the point of view of particular social perspectives. Drawing on Ásta Sveinsdottir’s “conferralist” account of social properties, I will suggest that membership in a variety of social categories membership is *conferred* upon individuals by agents and institutions. These “conferrers” assign category membership based on whether or not an individual seems to possess a given

property, from the perspective of the conferrer. This will allow us to get clearer on just how a given property of an individual could be causally sufficient to realize their membership in a given social category, at a context. Moreover, this framework will help us meet our second goal-- when conferral events are *successful*, individuals actually come to have the properties conferred on them.

3.1 Conferralism and social categories

In a recent series of writings, Ásta Sveinsdottir (2008, 2011, 2013, 2018) has advanced what she calls a “conferralist” account of the construction of human social categories. She notes that the basic metaphysical debate between social constructionists and non-constructionists is ultimately a reformulation of the Euthyphro problem: are people *men*, *women*, or *homosexuals* because they have an intrinsic category-specific property, or do they have their category-specific properties because they are believed to be *men*, *women*, and *homosexuals*? The social constructionist endorses the latter possibility: the existence and relative cohesiveness of a category like *woman* is in some sense dependent on our beliefs and social practices, not on any intrinsic feature of the individuals in the category. Consequently, the social constructionist owes us an account of *which* sorts of social practices give rise to these category-specific properties, *how* they manage to do so, and *what* sorts of properties ascriptions of *manhood*, *womanhood*, or *homosexuality* are tracking.

Ásta argues that social constructionists can meet these burdens by thinking of properties such as race and gender as *conferred properties*. A conferred property is any property which is dependent on our beliefs and practices quite generally, and in particular dependent on an individual, group, or institution with the power or standing to “confer” that property on an individual or object. For example, as foreshadowed in Chapter 2, “cool” is a paradigmatically conferred property. The

property of being cool entirely depends on there being people with dispositions to treat certain things as cool. Anything which has the property of being cool has had that status conferred on it by someone or something with the standing to do so. Indeed, it is difficult to make sense of what it is for something to be *cool* (or what the term “cool” means) without there being such attitudes.

The events by which certain properties are conferred need not be explicit. Sometimes conferral events do have this flavor of explicitness, as when a marriage official confers the property of *being married* on two individuals by virtue of saying “I now pronounce you married.” Yet conferral events are often tacit, or implicit as well, as when every attendee at a party simply comes to have the belief that a certain guest is *cool*. What is required for a conferral event to be successful is not that the conferral occur in a particular way, but that the conferrer have the authority or standing to make the conferral.

So conferred properties are metaphysically dependent on some conferral event, as well as the existence of conferrers of some sort. Yet they also depend on there being some underlying property that conferrers are attempting to *track* with their conferrals. Return to the example of baseball. The property of being a *strike* is a conferred property, and the relevant conferral event is a “strike” call made by an umpire. But the underlying properties that the umpire is attempting to *track* with his “strike” call is something like those trajectories of pitched balls that end up passing over home plate at a distance between a batter’s knees and chest. Call the properties that conferrers are trying to track via their conferrals *grounding properties*.¹⁰⁴

¹⁰⁴ Ásta has variously referred to these tracked properties as “grounding properties” (2008, 2011) and “base properties” (2018). I use the term “grounding property” here to distinguish the notion of a property that a conferral tracks from my use of the term “base property,” which refers to the property that is causally responsible for an individual’s membership in a social category at a context. Grounding properties may be identical with base properties in particular cases, but the two notions are conceptually distinct.

The baseball example makes clear that a conferral event might be *successful* without it being the case that the subject of the conferral actually has the grounding property that the conferral tracks. That is, a pitch can be a *strike* even if it does not have a trajectory which passes through the strike zone. The relation between the grounding property being tracked and the actual conferral of the property is “merely epistemic,” not causal. Conferrals are never “caused” directly by the presence of a given grounding property, although subjects might *believe* that they cause the conferral.

Note that it might not be explicit to a conferrer just *which* property is being tracked by a given conferral event. As before, a conferrer may think that their conferral of a property (e.g. *being a terrorist*) is justified by an individual’s having property P (being violent), when what *really* is guiding the conferral is the fact that an individual has property Q (being Muslim). Moreover, a conferral event might be tracking a *cluster* of grounding properties, none of which is individually epistemically sufficient for the conferral. For instance, in many contexts, conferrals of gender statuses such as *man* and *woman* do not only track the perceived or imagined possession of male or female genitalia, but the possession of a whole range of features including secondary sex characteristics, mannerisms, bodily comportment, dress, and the like.

To summarize, Ásta’s conferralist account answers five related questions:

What is conferred?: A social property P, e.g. being of gender G, or race R, being a strike, being a skinhead, being a hipster, being a professor...

Who does the conferring?: A person or entity with relevant authority or standing to confer the property.

What does the conferring?: An explicit conferral event, such as a speech act (in public, institutional contexts), or an implicit or explicit judgment (in more private contexts), e.g. “strike!” call in baseball, beliefs and ascriptions that hold that a certain person is a *man* or *woman*.

When is the conferral successful?: Particular contexts and appropriate circumstances, e.g. when “I pronounce you married” is felicitous, when “strike” is felicitously uttered by an umpire

What is being tracked by the conferral?: The “grounding property” that conferrers are attempting to track via ascriptions of the conferred property, e.g. the physical trajectories of baseballs in the case of “strike,” genitalia and/or self-identification in the case of “woman,” levels of income in the case of “hipster,” political beliefs or dress in the case of “skinhead,” etc. Which grounding properties are being tracked by a particular conferral can be a highly context-dependent matter.

Conferralism is friendly to our purposes for three reasons. First, the conferral of a single property might track different grounding properties in different contexts. As Ásta explains in applying her theory to the metaphysics of sex and gender,

My suggestions as to how gender is conferred makes gender out to be highly context-dependent. Not only is it deeply context-dependent when it comes to historical periods and geographic locations, but the same geographical location and time period can allow for radically different contexts, so that a person may count as a woman (or man or what have you) in some contexts and not others. This is because different properties are being tracked in different contexts: in some contexts it is perceived role in biological reproduction, in others it is role in societal organization of various kinds, sexual engagement, presentation of the body, in the preparation of food at family gatherings, etc. (Ásta 2011: 60-61)

This sort of context-dependence is precisely what we should predict of the properties corresponding to SRCs. In some contexts, *criminal* tracks objective patterns of law-breaking, and in others it tracks racialized properties, in still others it tracks both sorts of properties, and in others it tracks different properties altogether. Thinking of *criminal* as a conferred status allows us to account for this variation.

Second, the conferralist account allows us to be a bit more committal about what conditions must be fulfilled in order for a property of an individual to “realize” their membership in a social category. We can now understand a realizer property as a property of an individual which serves as

the basis of a successful conferral event. Realizer properties are just those properties which 1) can serve as the “grounding properties” that conferral events track, 2) where such conferral events are “successful” and causally sufficient to confer social statuses on individuals.

Finally, the conferralist account shows us that racialization processes can be metaphysically significant phenomena. For instance, assuming that the sorts of dispositions that racialized discourse cultivates (Chapters 5 and 6) are bound up with our dispositions to track certain racialized properties with our conferrals, these conferral practices will *actually make it the case* that individuals and groups who bear those racialized properties take on the conferred status. In the case of categories like *criminal*, *terrorist*, and *immigrant*, this means that those with particular racialized properties will be over-represented among the category. Yet it also means that the basic structure of the categories *themselves* changes insofar as racialization introduces new ways for individuals to realize membership in the category.

3.2 Conferralism and SRCs

The social categories we are interested in can largely be considered conferred statuses, and the properties that realize membership in those categories are the “grounding properties” that are tracked by conferral events. A general conferralist account of SRCs will thus be something like the following (adapted from Ásta 2011: 61):

Conferred Property: being a *criminal*, *terrorist*, *welfare recipient*, *thug*, etc.

Who (conferrers): the conferring institutions or subjects S in the particular context

What (conferral event): the perception of the conferring institutions or subjects S that the person have the grounding property P

When (conferral event): in context C

Tracking: the grounding property P that explains the conferral

A given conferral of the status *criminal*, *terrorist*, *welfare recipient*, or *thug* will be racialized to the extent that the conferral tracks a racialized grounding property P (e.g. *non-White*, *Muslim*, *Black*, etc.)

Whether any such conferral tracks P will depend on features of the context.¹⁰⁵

To see how this works in more detail, let's return to the racialization of the category *criminal* through the lens of Weaver's analysis. There we saw that the conservative strategy in the wake of incontrovertible Civil Rights gains was to focus the criminal apparatus on particular types of law-breaking such as protesting and rioting, and to symbolically connect those patterns of law-breaking to the broader Civil Rights struggle. I suggested that during this "frontlash" process, the nature of *criminality* somehow changed. We gave metaphysical content to this idea through our discussion of "tracked properties" and "realizer properties:" in particular, the social category *criminal* began to more aggressively track the racialized properties of individuals, and these same properties became more likely to realize an individual's criminal status.

We can now understand this process as largely underwritten by a cascade of conferral events represented at many different levels of social structure. We can somewhat artificially separate these into *political* levels, *institutional* levels, and *interpersonal* levels. Central to Weaver's story is that racializing crime and punishment was very much an *elite-led* movement to resecure White dominance. Her account suggests that we ought to locate the most important conferral events involving *criminal* at the *political* level populated by elected officials. These conferrals will have *institutional-level* ramifications for how sub-governmental agencies such as police forces and judiciaries will "frame" the *criminal* category. And finally, both of these levels will have ramifications for how *criminality* and

¹⁰⁵ Note that on this gloss racialized conferrals are not *ipso facto* problematic. Rather, they are problematic when they track racialized properties *unjustly*, or to the exclusion of candidate grounding properties that the conferral is "supposed" to track.

its related properties will be conferred within the thought and talk of normal citizens at the *interpersonal* level. In what follows, I focus on the political level, with the understanding that structurally similar conferral events will occur at other levels of social interaction (some dynamics of which have been traced in earlier chapters):

Political level conferrals

Conferred Property: being a criminal

Conferrers: elected officials, lawmakers

Conferral events: speech acts reflecting lawmakers' explicit decision to redefine Civil Rights struggle in terms of crime (rather than politics)

Properties tracked by conferral: rioting, looting, involvement in Civil Rights struggle

Weaver's paper offers a number of examples of speech acts from lawmakers that, on my understanding, count as conferral events for the property *criminal*. In May 1961, Mississippi Democrat Senator John Stennis explicitly claimed that the Civil Rights movement-- particularly the work of the "freedom riders"-- was a matter for the criminal code, not for "the political arena:"

As to the so-called freedom riders, it is well known that this trip originated far from the South, and had as its announced purpose a trip through several States where local statutes would be violated Instead of being encouraged by groups, and cast into the political arena, to be made a political issue, they should be made a violation of the criminal statutes. (quoted in Weaver 2007: 241)

Similarly, Barry Goldwater explicitly connected the Civil Rights struggle to criminality in a number of 1964 speeches:

[C]hoose the way of this present Administration and you have the way of mobs in the street, restrained only by the plea that they [referring to black militants and civil rights demonstrators] wait until after election time to ignite violence once again. (quoted in Weaver 2007: 242-243)

These conferral events effectively treat involvement in Civil Rights struggle-- and particularly Black involvement in such struggles-- as a realizer property for membership in the category *criminal*.

Equally important to explicit conferral events linking the Civil Rights struggle with criminality was the attempt to disarticulate the Civil Rights struggle from legitimate grievances. For instance, Democratic Senator Robert Byrd explicitly attacked the notion that Civil Rights unrest was a predictable and defensible reaction to deplorable inner city conditions:

So, Mr. President, ... [p]overty neither provides a license for laziness nor for lawlessness. We can take the people out of the slums, but we cannot take the slums out of the people. Wherever some people go, the ratholes will follow All the housing and all the welfare programs conceivable, will not stop the riots or do away with the slums. (Weaver 2007: 248-249)

Similarly, there was a concerted effort to downplay the fact that Civil Rights unrest was often catalyzed by instances of police brutality:

According to one “top authority,” Los Angeles Police Chief William Parker, the “police brutality” allegations were a “terribly vicious canard which is used to conceal Negro criminality ... to try to find someone else to blame for their crimes.” In his statement before the Crime Commission, J. Edgar Hoover, the highest ranked U.S. law enforcement official, dismissed police brutality as part of a communist campaign, stating that the “net effect of the charge of ‘police brutality’ is to provoke and encourage mob action and violence by developing contempt for constituted authority.” (Weaver 2007: 248)

On the vocabulary we have assembled here, these moves effectively make it the case that being involved in Civil Rights struggle could not serve as a realizer property in some *other*, non-criminal category. Under dominant political rhetoric, the fact that a group of people were *rioting*, for instance, could only realize membership in the category *criminal*, and not, say, in the category *activist*.

4. Conclusion: Racialization as a Metaphysical and Linguistic Phenomenon

The purpose of this chapter has been to illustrate some possible metaphysical upshots of my theory of racialized discourse. We have given content to notion of a “racialized” category, and articulated some of the mechanisms by which social categories become racialized. The core move has been to treat racialization as a process or strategy which imbues the social world with racial features. At this level of abstraction, racialized discourse is one such mechanism of racialization insofar as it engages cognitive and practical dispositions to represent and act on the social environment in ways that construct racialized categories.

Note that the picture of the metaphysical structure of racialized categories and our earlier account of the semantic structure of racialized terms represent two sides of the same coin. On the semantic side of the coin, Chapter 4 argued that we can think of racialized terms as having potentially race-specific *property thresholds*. The value of those thresholds are set by the context, as well as the nature of the social perspectives at play therein. Individuals are ruled into the extension of racialized terms to the extent that they are seen as having some cluster of properties central to these thresholds. On the metaphysical side of the coin, we find that membership in a “shifty” racialized category is a matter of having that status *conferred* by an institution, individual, or community with the standing to do so. In the same way that lexical items may be matched with a variety of possible property thresholds and property clusters, conferral events may track a variety of grounding properties.

We have now come full circle. We began with a discussion of how *semantic* racialized discourse represents the world, and how it distributes those representations in concrete linguistic practices. Among other things, those representations essentialize their targets, and naturalize racial differences. Along the way, we discovered that racialized discourse doesn’t purport merely to say

something about the way the world “is,” but says something about how the world *ought* to be. Racialized discourse increases tolerance for social hierarchy, and can be thought-guiding and action-guiding in ways that lead its interpreters to strive to uphold the racial status quo. In some cases, racialized discourse serves as a tool for organizing the racial domain. Under the conditions of racial domination, racialized terms and concepts are applied to specific populations. What we have tried to show in the last part of this argument is that given certain conditions, and within the idiolect of the dominant racialized vocabulary, those populations can actually come to have the properties those terms and concepts ascribe. Blacks become criminal. Muslims become terrorists. Immigrants become animals.

As we noted in the introduction, these two domains of racialized discourse are co-dependent. Each domain needs the other in order to properly function. The transmission of racialized representations through language creates racial realities. Racial reality, carved as it is in the image of racialized discourse, then repays language in kind, providing the conditions under which racialized discourse can seem truthful, reasonable, and normal.

Yet perhaps knowing how the beast works will give us clues about how to slay it. For instance, if the conferralist approach to the construction of racialized categories is on the right track, there are a number of points at which such construction could be *resisted*. Activists of various stripes might challenge the *authority* of those bodies which do the conferring, such as in the context of protests against policing and incarceration. They may shift the *context* in which such conferral events take place, in order to make such conferrals unsuccessful-- as in when resistant communities create shared norms which block the success of attempted conferrals of “riot” or “thug.” They might challenge the legitimacy of a *type* of conferral event, as when activists push back against the use of

concepts like “illegal alien” or “immigrant caravan.” And finally, they might devote their efforts to changing the *grounding properties* that are tracked by conferral events, as when activists and policymakers undertake direct action to alter the life chances of those living in racially marginalized communities. At that point the descriptive-explanatory theory of racialized discourse we have pursued here will meet up with genuinely normative considerations and practical problems. And then the real work will start.

Appendix. Against Semantic-level Perspectivalism for Racialized Terms

In Chapter 2, we saw that social perspectives might play one of two roles in determining linguistic meaning, corresponding to two different senses in which a term might be “perspectival.” First, social perspectives might be part of the semantic value of a perspectival term. Call this view *semantic-level perspectivalism*. Second, social perspectives might explain why a perspectival term receives the semantic value it does at a context. Call this view *metasemantic-level perspectivalism*. As we saw in Chapters 2-4, my sympathies lie with metasemantic-level perspectivalism. This appendix briefly outlines the basic proposal of semantic-level perspectivalism, and provides some reasons for thinking that it is unworkable when it comes to understanding the semantics of racialized terms.

1. Semantic-level Perspectivalism

In recent work (O’Donnell 2017), I argued that a wide variety of racialized terms-- particularly gradable adjectives often used to express racial stereotypes such as “criminal,” “lazy,” “violent,” “entitled,” etc.-- are semantic-level perspectival. In a manner parallel to Lasersohn’s (2005) and Glanzberg’s (2007, ms) treatment of predicates of personal taste, I argued that SPPs (and *a fortiori* racialized terms) bring to the context of use a contextual parameter, **RP** which takes an argument denoting a *relevant perspective*. In the same way that the **E** parameter is unique to predicates of personal taste (see discussion in Chapter 4, Section 4.1), the RP parameter is unique to socially perspectival predicates. As a result, the meaning of a particular SPP can be informally defined as follows:

[P] = Property P-according-to-RP

The semantic clauses of SPPs were thus understood to be directly isomorphic to Glanzberg's two-parameter analysis of predicates of personal taste:

[cool] = exceeds-cool-standard-S-according-to-RP
[violent] = exceeds-violence-standard-S-according-to-RP

And in the case of count nouns such as “terrorist,” “foreigner,” and the like, we should expect to find only one contextual parameter:

[terrorist] = terrorist-according-to-RP
[foreigner] = foreigner-according-to-RP
[immigrant] = immigrant-according-to-RP
[criminal] = criminal-according-to-RP
[thug] = thug-according-to-RP

Yet unlike the standard parameter, which only appears in the logical form of gradable adjectives, and the experiencer parameter, which only appears in the logical form of adjectival predicates of personal taste, the RP parameter appears in a variety of morphosyntactic guises. The semantic value of these terms is “incomplete at a context” until a relevant perspective is assigned.

According to this story, the reason that an RP belongs in the semantics is that SPPs describe perspective-dependent properties: in order for anything to count as a criminal, there needs to be some perspective from which a given person is judged to be criminal. In other words, one is only a criminal *for* some code, social structure, or judicial system. Indeed, one source of linguistic evidence for the RP parameter is that terms like “criminal” and “terrorist” can be complemented by *to/for/according to* adjuncts, as in “Osama bin Laden was a terrorist according to everyone in the

United States [but was a freedom-fighter according to others]” or “Pot-smokers are criminal to Turkish police, but not not to Dutch police.”¹⁰⁶

2. Different Types of Perspectives

One of the benefits of the RP analysis is that it nicely handles cases in which the relevant perspective is a perspective shared by an individual (as in the case of “strike,”) as well as cases in which the relevant perspective is a perspective shared by a *class* of agents, or even a *generic* set of possible perspectives distributed across the social environment.

Let me explain. Some uses of so-called predicates of personal taste do not report on what an object or experience is like *to* a particular experiencer, but seem to make a claim about how things are in the real world, independently of the perspective of any *particular* experiencer, but dependent on the perspective of possible *generic* experiencers in the linguistic community. To see this, consider two different ways of interpreting 1).

1) Beyoncé is not cool.

What is the relevant experiencer class? Well, if 1) could be true in a context in which I am understood to be reporting on my own opinions about Beyoncé. But if the use of “cool” here is supposed to pick out generic experiencers, I would be saying something false if I utter 1) in a community in which Beyoncé is widely and generically regarded as cool.

This ostensibly dual function of “cool” seems to give us reason to think that the first pair of sentences below describes a disagreement, while the second pair does not. Assuming the speaker is

¹⁰⁶ We’ll look more closely at this “evidence” shortly, however.

being truthful, a retraction or clarification of A's position following the first pair of sentences seems entirely natural, but not the second pair.

A: Beyoncé is not cool [generically].
B: What are you talking about? Of course she is!
A: Ok, well she's not cool *to me*. I have different tastes.

A: Beyoncé is not cool [to me].
B: What are you talking about? Of course she is!
A: ?? You're right. I lied before.

In the first set, the speaker backs down from the strong claim that Beyoncé “lacks coolness,” so to speak, in order to report on her personal dispositions about what she finds cool. In the second set, the speaker begins with a report on her dispositions, and her interlocutor either 1) misunderstands that the speaker is using “cool” in this private way and challenges her, 2) reports on his own dispositions regarding coolness, thus failing to disagree with the speaker, or 3) understands that the use is private, but maintains that the speaker misunderstands what the proper extension of “cool” is. In all three cases, the speaker's retraction is infelicitous.¹⁰⁷

Now, if “cool” is a predicate of personal taste, we need some experiencer class to serve as the value of E. Yet the idea that an entire culture could be the most relevant experiencer class for the truth-conditions of the predicate sounds metaphysically odd. After all, *cultures* aren't generally the sorts of things that have experiences, and they certainly aren't the sorts of things that *report* on their experiences. There are thus at least *prima facie* reasons to think that cultures or generic perspectives can't serve as the value of the E parameter.

¹⁰⁷ In the third case, perhaps B's utterance could lead to a negotiation over which perspective on 'cool' is the most appropriate one, and perhaps A might eventually retract her statement after endorsing B's conception of 'cool.' However, immediate retraction is odd, to say the least. (And even when it occurs, this situation can plausibly be read as feigned endorsement in order to curry favor with B rather than as a true change of semantic heart.)

On the other hand, it does seem that the two possible interpretations of “cool” in 1) represent a difference worth marking, and we need some technology for marking it. My claim was that the RP parameter might be saturated with *either* individualistic, cognitive perspectives *or* with shared, distributed perspectives “out there” in the discourse community at large. It’s worth treating this point in a bit more detail.

Sometimes culture-dependent *facts* about what is cool, sexy, beautiful, etc. are instituted along with dominant understandings of what these predicates like “cool” purport to pick out. Marketing tactics provide us with a ready example of this sort of institution. Assume that after months of painstaking research, a company markets a product to a coveted demographic. The major selling point of this product is that it is “cool” to a large number of consumers in that demographic. Roughly, there are two ways that the company can get these potential consumers to judge the product to be “cool.” First, they could monitor the cool-judging dispositions of their target demographic, and, using this data, try to sell the members of that demographic a product that they are likely to think is “cool.” Second, and more ambitiously, they could indirectly *cultivate* or *manipulate* the cool-distinguishing dispositions of their target demographic in a way that ensures that they will find the product “cool.”

Call those responsible for tracking or shaping cool-distinguishing dispositions *cool-producers*, and those whose dispositions are so shaped *cool-consumers*. Once this loop between cool-producers and cool-consumers is established, things get metaphysical (and profitable.) Cool-producers carefully monitor and cultivate the cool-distinguishing dispositions of cool-consumers, and when all goes well, cool-consumers start judging the things that cool-producers produce as “cool.” Cool-producers and cool-consumers thus participate in the establishing of a *kind* consisting of cool objects and

properties. And when a certain mode of cool-production and cool-consumption becomes *dominant* in a society or culture, this opens up the possibility for certain dissenters to be just *wrong* about what is cool for that society or culture. This does not mean that individual dissenters are not entitled to their opinion, but it does mean that individual dissenters can speak falsely when they are attempting to speak about what is cool *from the perspective of the culture*.

Aesthetic sensibilities in particular can cluster in ways that produce a social-normative *consensus* about art, pleasure, and politics. We often find ourselves, especially in our more unreflective moments, both participating in and parroting this consensus in ways that establish what may seem like “the facts” about what is aesthetically valuable, pleasurable, shocking, etc. In describing the phenomenology of everyday life, Martin Heidegger nicely captures the ways in which these individual dispositions and behaviors conform in ways that yield judgments that are widely accepted at the level of social structures and cultures:

We take pleasure and enjoy ourselves as *the one* [*das Man*] takes pleasure; we read, see, and judge about literature and art as *the one* sees and judges; likewise, we shrink back from the “great mass” as *the one* shrinks back; we find shocking what *the one* finds shocking. “The one,” which is nothing definite, and which all are, though not as the sum, prescribes the kind of Being of everydayness. (Heidegger 1962: 164).

Heidegger thus emphasizes a distinction between the lived, embodied perspective of the individual, and then the abstract, amorphous perspective of *the one*, the faceless, generic (“nothing definite”), emergent (“all are, though not as the sum”) arbiter of what is pleasurable, aesthetically valuable, shocking, etc. Both of these perspectival vantage points (and many more besides) are points of view from which one can think, speak, and reason about these phenomena.

Back to semantics. Pairing this hypothesis with the usual semantics for gradable adjectives discussed in Chapter 2, the meaning of ‘criminal’ in context thus receives the following analysis:

2) [[criminal]]^c= **degree-criminality-for-RP**

Once we introduce the RP parameter, we can revisit our semantics for “Blacks are criminal.”

- 3) a. Blacks are criminal.
- b. The degree of criminality possessed by Blacks as a kind according to a relevant perspective is greater than the standard degree of criminality according to that perspective.
- c. $\text{GEN criminal}_{\text{RP}}(\text{Blacks}) > s(\text{criminal}_{\text{RP}})$

Again, in providing these truth conditions, we have not yet specified the *particular* perspective which ‘criminal’ picks out. In the contexts we’re interested in, those perspectives are selected by facts about racialized social structure, just as baseball perspectives are selected by facts about how baseball is conducted. This semantics makes clear that RPs play a role both in determining the extension of SPPs and in determining the *standard* in virtue of which which the properties they describe are associated with race.

This view is defended at more length in O’Donnell (2017). Yet the hope behind the strategy was that this semantic treatment would provide us with a ready account of why terms like “thug” and “terrorist” are subject to race-related shifts in meaning-standards. For instance, the semantics (and not the meta-semantics, as I ended up arguing in Chapter 4) would effectively explain the fact that Muslims (and not Whites) are routinely described as terrorists by appeal to the fact that the dominant social perspective on “terrorism” treats Muslim identity (and not White identity) as a central feature.

However, while I believe that semantic-level analysis may still be viable for some SPPs, I have come to believe that a global treatment of SPPs-- and in particular racialized terms-- as semantic-level perspectival is confronted with serious problems. I'll briefly discuss three of these.

3. Three problems for semantic-level perspectivalism

3.1 *Is an RP parameter necessary?*

If the RP analysis of racialized terms is correct, a racialized term does not secure a semantic value in context until it is assigned a perspective. Assuming that “securing a semantic value” for an expression entails having a hearer assign an interpretation to that expression, this entails that “thug,” “terrorist,” “criminal,” and the like do not make their contribution to truth-conditional content without having the RP parameter saturated. Yet intuitively, these terms *do* secure a referent, even if the hearer does not interpret the term from any particular perspective. A competent hearer of a sentence such as “James is a thug” easily grasps a set of conditions under which the sentence is true-- roughly, that James is a brutish or violent person-- *regardless* if her interpretation of “thug” is racialized or not.

3.2 *Lack of evidence from syntax*

Next, note that a semantic-level theory cannot avail itself of the syntactic arguments usually used to demonstrate the perspectival nature of predicates of personal taste (see Chapter 4, Section 4.1.) While predicates of personal taste can easily accommodate “to” or “for” adjuncts,¹⁰⁸ the following locutions sound decidedly strange:

¹⁰⁸ As Stephenson (2007) points out, which adjunct can be selected seems to depend on the meaning of the term itself.

4) He is a thug to me/for me

5) Bin-Laden was a terrorist to me/for me

Of course, there is a reading of each of these expressions which sounds natural: when “to-X” and “for-X” are thought of as elliptical for “according to-X.” Yet we should be careful of according any evidential weight to this reading. After all, if the “according-to-X” adjunct is sufficient for rendering a term semantic-level perspectival, the proposal rapidly overgeneralizes to wide swaths of linguistic items that do not seem related by any interesting semantic or syntactic principles:

6) That is a dog according to me/you/them.

7) Those leaves are green according to me.

8) Binky is here according to me.

9) Samuel Clemens is Mark Twain according to me.

Each of these sentences has a natural, acceptable reading, yet none of them convincingly demonstrates that the “according to” adjunct makes explicit hidden structure in their adjacent lexical items.

But even if a bullet-biting defender of semantic-level perspectivalism *were* to take this as evidence that most or all terms are perspectival, this would seem to trivialize the entire proposal. After all, the motivation for semantic-level perspectivalism is to isolate a class of terms which describe those objects whose properties are somehow determined by the perspectives of those who classify them. These motivations are in turn both metaphysical and semantic. On the metaphysical level, we may want a principled way of distinguishing *interactive* kinds, or kinds which are somehow responsive to the way they are classified, and *indifferent* kinds, or kinds which do not seem responsive in this way (Hacking 1999.) Correlatively, on the semantic level, we want a way of distinguishing, in

an informative way, the general semantic profile of sentences like “Snakes are slimy” on the one hand and “Drugs are cool” on the other. Yet both of these motivations are undercut if we trivialize the thesis of semantic-level perspectivalism.

3.3 The Binding Criterion

Finally, and perhaps most problematically for a semantic-level theory, it does not seem that many paradigmatic racialized terms contain an implicit argument of any kind. One (very strong) source of evidence for this is the fact that the interpretation assigned to racialized terms does not seem to co-vary with the scope of a quantifier domain. In two seminal papers, Stanley (2000) and Stanley and Szabo (2000) propose a test for determining whether the interpretation of a given “contextual ingredient” *C* in a sentence is due to “saturation”-- roughly, to the presence of a *variable* (or argument) represented in the syntax of *C*-- or to “free enrichment”-- roughly, to the presence of “unarticulated constituents” which are not represented in the syntax, but are supplied by spontaneous inferences that interpreters bring to the interpretation process in order to add some contextual information which is not contained in the “literal meaning” of *C*. The test, which the authors call the *Binding Criterion*, is as follows:

Binding Criterion.

A contextual ingredient in the interpretation of a sentence *S* results from saturation if it can be “bound,” that is, it can be made to vary with the values introduced by some operator prefixed to *S*.

Let’s consider what results the binding criterion gives us in the case of quantifiers. Suppose you and I are looking for something to drink, and we come across an overturned beer truck. After looking

around, I say, disappointedly:

10) Every bottle is empty.

Naturally, you do not take me to be saying that every bottle *in the universe* is empty. Rather, you take me to be saying that every bottle in the salient domain--on the truck, in the street, etc.-- is empty. More technically put, you automatically interpret the meaning of the quantified expression “every bottle” to *restrict* the domain to some salient domain x :

11) Every bottle [in x] is empty.

The question is, what allows you to assign the correct domain restriction? Again, for Stanley and Szabo, there are two options: either you are assigning a semantic interpretation to a variable located somewhere in the syntax of 10), or you are responding to the presence of a non-syntactic “unarticulated constituent” in 10) and inferring the correct domain restriction through a process of “free enrichment.” Importantly, if the latter is the case, 10) strictly speaking expresses a *false* proposition-- that every bottle *in the universe* is empty-- and the utterance is judged true only because you spontaneously add the missing contextual information. Which theory is correct? Enter the Binding Criterion. According to Stanley and Szabo, only one of these theories can account for the most natural readings of sentences involving *multiple* quantifier domains:

12) In every room in John’s house, every bottle is in the corner.

The most natural reading of 12) is something like the following:

13) For every x such that x is a room in John’s house, every bottle in x is in the corner of x .

That is, the second quantifier phrase (“every bottle”) automatically restricts to the domain introduced by the first quantifier phrase (“every room.”) The sentence is *not* naturally interpreted as

follows:

- 14) For every x such that x is a room in John's house, every bottle [in the universe] is in the corner of x .

Since the interpretation of the domain of the second quantifier co-varies with the interpretation of the domain of the first quantifier, the Binding Criterion rules that the natural interpretation of this sentence is a result of “saturation,” not “free enrichment.” Moreover, notice that a proponent of the “free enrichment” theory has trouble explaining how we get reading 13) and not reading 14). If there are no syntax-driven constraints on the interpretation of the scope of the quantifier domains, how is it that competent speakers of the language converge on 13)? In fact, reading 14) seems not just false, but unintelligible.¹⁰⁹

The Binding Criterion also determines that sentences like the following include hidden variables:

- 15) Every species has members that are small.

The most natural interpretation of this sentence is that every species has members who are small *by the standards of the species*. A far less natural reading is that every species has members that are small *simpliciter*. (For instance, it would be inappropriate to say that 15) is false on the grounds that *whales* are not small.) Since the interpretation of “small” varies with the interpretation of the quantifier, this is evidence that the quantifier binds an implicit argument in the syntax of “small.”

Yet again, it's not clear that we get a restricted reading for many racialized terms when we

¹⁰⁹ The possible responses for the free enrichment theory don't stop here, of course. Defenders of that view could distinguish sentences which allow “optional” vs. “mandatory” types of free enrichment, and propose their own non-syntactic criteria for determining which linguistic cases are which. See Recanati (2004) for a response along these lines. Yet since I only introduce the Binding Criterion in order to apply it to the case of possible hidden structure in perspectival and racialized terms, I leave this particular debate behind.

embed them in sentences of this type. According to semantic-level perspectivalism, racialized terms should include a contextual parameter which takes a relevant perspective, and the term gets its racialized interpretation by having its semantic value computed from that semantic perspective (analogously to the way that the semantic values of predicates of personal taste are computed.) In order for something like this hypothesis to be correct, there needs to be something like a variable in syntax which corresponds to the RP parameter. Yet consider the following sentences:

16) Every race has members who are violent.

17) Every race has members who are thugs.

18) Every race has members who are terrorists.

I myself do not get the reading according to which each race has members who are violent *by standards of that race*. It seems even clearer to me that 17) and 18) only receive readings according to which every race has members who are thugs or terrorists *simpliciter*.

Bibliography

Alexander, M., 2012. *The New Jim Crow*. New York: The Free Press.

Allport, G. W., 1958. *The Nature of Prejudice*. New York: Doubleday.

Angwin, J., J. Larson, S. Mattu, and L. Kirchner., 2016. Machine bias. *ProPublica*. May 23, 2016

Asher, N. and Morreau, M., 1995. What some generic sentences mean. In *The Generic Book*, edited by Greg Carlson and Francis J. Pelletier, 300-339. Chicago: Chicago University Press.

Ásta Sveinsdóttir., 2008. Essentiality conferred. *Philosophical Studies*, 140(1), pp.135-148.

Ásta., 2011. "The metaphysics of sex and gender." In *Feminist metaphysics* (pp. 47-65). Springer, Dordrecht.

Ásta., 2013. The social construction of human kinds. *Hypatia*, 28(4), pp.716-732.

Ásta., 2018. *Categories We Live By*. Oxford: Oxford University Press.

Barsalou, L. and Sewell, D.R., 1984. Constructing representations of categories from different points of view. Emory university lab report.

Barsalou, L., 1987. The instability of graded structure: Implications for the nature of concepts. *Concepts and conceptual development: Ecological and intellectual factors in categorization*.

Bastian, B. and Haslam, N., 2006. Psychological essentialism and stereotype endorsement. *Journal of Experimental Social Psychology*, 42(2), pp.228-235.

Bastian, B. and Haslam, N., 2007. Psychological essentialism and attention allocation: Preferences for stereotype-consistent versus stereotype-inconsistent information. *The Journal of Social Psychology*, 147(5), pp.531-541.

Bastian, B. and Haslam, N., 2008. Immigration from the perspective of hosts and immigrants: Roles of psychological essentialism and social identity. *Asian Journal of Social Psychology*, 11(2), pp.127-140.

Bertrand, M. and Mullainathan, S., 2004. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American economic review*, 94(4), pp.991-1013.

Billings, A.C. and Eastman, S.T., 2002. Selective representation of gender, ethnicity, and nationality in American television coverage of the 2000 Summer Olympics. *International review for the sociology of sport*, 37(3-4), pp.351-370.

Blackmon, D. A. 2009., *Slavery by Another Name: The re-enslavement of black Americans from the Civil War to World War II*. Anchor.

Blauner, B. 1992., Talking past each other: Black and white languages of race. *American Prospect*, 10, pp.55-64.

Blow, C. 2014., Crime, bias, and statistics. New York Times, September 7.
<http://www.nytimes.com/2014/09/08/opinion/charles-blow-crime-bias-and-statistics.html>.

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

Brandone, A.C., A. Cimpian, S. J. Leslie, and S. A. Gelman., 2012. Do lions have manes? For children, generics are about kinds rather than quantities. *Child Development* 83 (2): 423–433.

Bryc, K., E. Durand, J. M. Macpherson, D. Reich, and J. M. Mountain., 2015. The genetic ancestry of African Americans, Latinos, and European Americans across the United States.” *The American Journal of Human Genetics* 96: 37–53.

Burgess, A. and Sherman, B., eds., 2014. *Metasemantics: New Essays on the Foundations of Meaning*. Oxford University Press, USA.

Camp, E., 2013. “Slurring Perspectives.” *Analytic Philosophy* 54 (3): 330–349, 2013.

Cappelen, H. and Lepore, E., 2008. *Insensitive semantics: A defense of semantic minimalism and speech act pluralism*. John Wiley & Sons.

Carlson, G., 1977. *Reference to Kinds in English*. Doctoral Dissertation, University of Massachusetts, Amherst.

Carlson, G. and Pelletier, F.J., eds. 1995. *The Generic Book*. Chicago: Chicago University Press.

Chierchia, G., 1995. Individual-level predicates as Inherent Generics. In *The Generic Book*, edited by Greg Carlson and Francis J. Pelletier, 176–223. Chicago, IL: Chicago University Press.

Cimpian, A. and Markman, E.M., 2009. Information learned from generic language becomes central to children’s biological concepts: Evidence from their open-ended explanations. *Cognition*, 113(1), pp.14-25.

Cimpian, A., Mu, Y. and Erickson, L.C., 2012. Who is good at this game? Linking an activity to a social category undermines children’s achievement. *Psychological science*, 23(5), pp.533-541.

Cohen, A., *Think Generic!: The Meaning and Use of Generic Sentences*. Stanford: CSLI Publications, Center for the Study of Language and Information, 1999.

Condit, C.M., Parrott, R.L., Harris, T.M., Lynch, J. and Dubriwny, T., 2004. The role of “genetics” in popular understandings of race in the United States. *Public Understanding of Science*, 13(3), pp.249-272.

Correll, J., B. Park, C. M. Judd, and B. Wittenbrink., 2002. The police officer’s dilemma: Using ethnicity to disambiguate potentially threatening individuals.” *Journal of Personality and Social Psychology* 83 (6): 1314–1329.

Correll, J.B., Park, C. M. Judd, and B. Wittenbrink., 2007. The influence of stereotypes on decisions to shoot.” *European Journal of Social Psychology* 37: 1102–1117.

Coryn, C.L., Beale, J.M. and Myers, K.M., 2004. Response to September 11: Anxiety, patriotism, and prejudice in the aftermath of terror. *Current Research in Social Psychology*, 9(12), pp.165-183.

Craig, M.A. and Richeson, J.A., 2014. 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), pp.1189-1197.

Craig, M.A. and Richeson, J.A., 2014. More diverse yet less tolerant? How the increasingly diverse racial landscape affects white Americans’ racial attitudes. *Personality and Social Psychology Bulletin*, 40(6), pp.750-761.

- Crawford, K., 2016. Artificial intelligence's white guy problem. *The New York Times*, 25.
- Curry, T. J., 2017. *The Man-Not: Race, Class, Genre, and the Dilemmas of Black Manhood*. Temple University Press.
- Davis, A.Y., 1998. Racialized punishment and prison abolition, in *The Angela Y. Davis Reader*. John Wiley & Sons.
- Declerck, R., 1986. "The Manifold Interpretations of Generic Sentences." *Lingua* 68: 149–188.
- Diaz-Leon, E., 2016. Woman as a politically significant term: A solution to the puzzle. *Hypatia*, 31(2), pp.245-258.
- Diaz Legaspe, J., 2018. Normalizing Slurs and Out-group Slurs: The Case of Referential Restriction. *Analytic Philosophy*, 59(2), pp.234-255.
- Dice, M., Trump speech to police officers-- law and order will return to streets of America. *Video*. Uploaded to YouTube July 28, 2017. Accessed January 25, 2019: <https://www.youtube.com/watch?v=i7gqHI-NsO0&t=1668s>
- Dilulio, J., November 27, 1995. The coming of the super-predators. *The Weekly Standard*.
- Doris, J.M., 2002. *Lack of character: Personality and moral behavior*. Cambridge University Press.
- Dovidio, J.F., Brigham, J.C., Johnson, B.T. and Gaertner, S.L., 1996. Stereotyping, prejudice, and discrimination: Another look. *Stereotypes and stereotyping*, 276, p.319.
- Dummett, M., 1975. What is a theory of meaning? In S. Guttenplan (Ed.), *Mind and Language*, pp. 97–138. Oxford: Oxford University Press. Reprinted in Dummett (1993).
- Dummett, M., 1993. *The Seas of Language*. Oxford: Oxford University Press.
- Duncan, B.L., 1976. Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of blacks. *Journal of personality and social psychology*, 34(4), p.590.
- Eastman, T., and A.C.Billings, S., 2001. Biased voices of sports: Racial and gender stereotyping in college basketball announcing. *Howard Journal of Communication*, 12(4), pp.183-201.

- Fanon, F., 2008. *Black Skin, White Masks*. New York: Grove Press.
- Fischhoff, B., Gonzalez, R.M., Small, D.A. and Lerner, J.S., 2003. Judged terror risk and proximity to the World Trade Center. In *The risks of terrorism* (pp. 39-53). Springer, Boston, MA.
- Fiske, S.T., 1998. Stereotyping, prejudice, and discrimination. *The handbook of social psychology*, 2, pp. 357-411.
- Fiske, S.T., Cuddy, A.J., Glick, P. and Xu, J., 2018 (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. In *Social Cognition* (pp. 171-222). Routledge.
- Frank, J.D., 1982. Prenuclear-age leaders and the nuclear arms race. *American Journal of Orthopsychiatry*, 52(4), p.630.
- Gabrielson, Ryan, R. G. Jones, and E. Sagara., 2014. Deadly force, in black and white. *ProPublica*. Accessed October 10, 2017. <http://www.propublica.org/article/deadly-force-in-black-and-white>.
- Gallie, W. B., 1955–1956. Essentially contested concepts. *Proceedings of the Aristotelian Society New Series*. Vol. 56, 167–198.
- Gelman, S.A., 2003. *The Essential Child*. Cambridge: MIT Press.
- Geurts, B., 1985. Generics. *Journal of Semantics* 4: 247–255.
- Gilens, M., 1999. *Why Americans Hate Welfare: Race, media, and the politics of antipoverty policy*. University of Chicago Press.
- Gill, P. and Horgan, J., 2012. Tracing the Motivations and Antecedent Behaviors of Lone-Actor Terrorism. *International Center for the Study of Terrorism, Pennsylvania State University*.
- Gilliam, F., 1999. The “welfare queen” experiment: How viewers react to images of African–American mothers on welfare. *Nieman Reports of the Nieman Foundation for Journalism at Harvard University*. Vol. 53, 2. <http://niemanreports.org/articles/the-welfare-queen-experiment/>
- Glanzberg, M., 2007. Context, content, and relativism. *Philosophical Studies* 136: 1–29.
- Glanzberg, M., Unpublished manuscript. Not all contextual parameters are alike.

Glasgow, J., 2010. *A Theory of Race*. Routledge.

Goodin, R.E., 2006. *What's wrong with terrorism?*. Polity.

Gramlich, J., 2018. America's incarceration rate is at a two-decade low. *Pew Research Center*, May.

Retrieved November 24, 2018.

(<http://www.pewresearch.org/fact-tank/2018/05/02/americas-incarceration-rate-is-at-a-two-decade-low/>)

Griffith, A.M., 2018. Social construction and grounding. *Philosophy and Phenomenological Research*, 97(2), pp.393-409.

Hacking, I., 1999. *The Social Construction of What?*. Harvard university press.

Halberda, J., and Feigenson, L., 2008. Developmental change in the acuity of the “number sense”: The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults.” *Developmental Psychology* 44 (5): 1457–1465.

Hamilton, C.V. and Ture, K., 2011. *Black power: Politics of liberation in America*. Vintage.

Heidegger, M., 1962. *Being and Time*, translated by J. Macquarrie and E. Robinson. Harper and Row.

Heyman, G.D. and Gelman, S.A., 1999. The use of trait labels in making psychological inferences. *Child development*, 70(3), pp.604-619.

Hamilton, D.L. and Sherman, S.J., 1996. Perceiving persons and groups. *Psychological review*, 103(2), p.336.

Haney-Lopez, I., 2014. *Dog Whistle Politics*. New York: Oxford University Press.

Harman, G., 1999. Moral philosophy meets social psychology: Virtue ethics and the fundamental attribution error. In *Proceedings of the Aristotelian society* (pp. 315-331). Aristotelian Society.

Harrison, C.K., Lawrence, S.M. and Bukstein, S.J., 2011. White college students' explanations of white (and black) athletic performance: A qualitative investigation of white college students. *Sociology of Sport Journal*, 28(3), pp.347-361.

Haslam, N., 2000. Psychiatric categories as natural kinds: Essentialist thinking about mental disorder. *Social Research*, pp.1031-1058.

Haslam, N. and Levy, S.R., 2006. Essentialist beliefs about homosexuality: Structure and implications for prejudice. *Personality and Social Psychology Bulletin*, 32(4), pp.471-485.

Haslam, N., Rothschild, L. and Ernst, D., 2000. Essentialist beliefs about social categories. *British Journal of Social Psychology*, 39(1) pp.113-127.

Haslam, N., Rothschild, L. and Ernst, D., 2002. Are essentialist beliefs associated with prejudice?. *British Journal of Social Psychology*, 41(1), pp.87-100.

Haslanger, S., 1995. Ontology and social construction." *Philosophical Topics* 23 (2): 95–125.

Haslanger, S., 2000. Gender and race: (What) are they? (What) do we want them to be?. *Noûs*, 34(1), pp.31-55.

Haslanger, S., 2005. What are we talking about?: The semantics and politics of social kinds." *Hypatia*. 20 (4): 10–26.

Haslanger, S., 2007. "But mom, crop-tops are cute!" Social knowledge, social structure and ideology critique. *Philosophical Issues*, 17(1), pp.70-91.

Haslanger, S., 2011. Generics, ideology, and common ground. *Feminist Metaphysics*, edited by Charlotte Witt, 179–207. Springer-Verlag. Reprinted in Haslanger 2012.

Haslanger, S., 2012. *Resisting Reality: Social Construction and Social Critique*. Oxford University Press.

Haslanger, S., 2014. The normal, the natural and the good: Generics and ideology. *Politica & Società*, 3(3), pp.365-392.

Heck, R., 2001. Do demonstratives have senses? *Philosophers' Imprint*, 2.2., 1-33.

Hewstone, M., 1990. The "ultimate attribution error"? A review of the literature on intergroup causal attribution. *European Journal of Social Psychology*, 20(4), pp.311-335.

Hewstone, M., & Jaspers, J., 1982. Cross-cultural Interaction, Social Attribution and Inter-Group Relations!. *Cultures in Contact: Studies in Cross-Cultural Interaction*, Pergamon, Oxford.

Hirschfeld, L.A., 1998. *Race in the Making: Cognition, Culture, and the Child's Construction of Human Kinds*. MIT Press.

Hoicka, E., Saul, J., Prouten, E., Whitehead, L. and Sterken, R., 2018. Generics aren't special: Language signalling high proportions leads to essentializing for novel social kinds. <https://doi.org/10.31234/osf.io/xe6s>

Howard, G., 1/21/2014. Richard Sherman and the plight of the conquering Negro. *Deadspin*. Accessed January 26, 2019. <https://deadspin.com/richard-sherman-and-the-plight-of-the-conquering-negro-1505060117>

Huber, G.A. and Lapinski, J.S. 2006., The “race card” revisited: Assessing racial priming in policy contests. *American Journal of Political Science*, 50(2), pp.421-440.

Hurwitz, J. and Peffley, M., 1997. Public perceptions of race and crime: The role of racial stereotypes. *American journal of political science*, pp.375-401.

Hurwitz, J. and Peffley, M., 2005. 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), pp.99-112.

Hutchings, V.L., Walton Jr, H. and Benjamin, A., 2010. The impact of explicit racial cues on gender differences in support for confederate symbols and partisanship. *The Journal of Politics*, 72(4), pp.1175-1188.

Jayarathne, T.E., Gelman, S.A., Feldbaum, M., Sheldon, J.P., Petty, E.M. and Kardian, S.L., 2009. The perennial debate: Nature, nurture, or choice? Black and White Americans' explanations for individual differences. *Review of General Psychology*, 13(1), p.24.

Jefferson, T., and Franklin, B., 1776. Preamble to the American Declaration of Independence.

Jenkins, K., 2016. Amelioration and inclusion: Gender identity and the concept of woman. *Ethics*, 126(2), pp.394-421.

Jones, E.E. and Nisbett, R.E., 1971. *The Actor and the Observer: Divergent Perceptions of Behavior*. Morristown.

Jost, J.T., Banaji, M.R. and Nosek, B.A., 2004. A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political psychology*, 25(6), pp.881-919.

Kaplan, D., 1989. Demonstratives. In *Themes from Kaplan*, ed. Joseph Almog, John Perry, and Howard Wettstein, 481-566. New York: Oxford University Press.

Kay, A., Jimenez, M. C., & Jost, J. T., 2002. Sour grapes, sweet lemons, and the anticipatory rationalization of the status quo. *Personality and Social Psychology Bulletin*

Keil, F. C., 1989. *Concepts, Kinds, and Cognitive Development*. Cambridge, MA: MIT Press.

Kennedy, C. and McNally, L., 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language*, pp.345-381.

Kennedy, C., 2007. Vagueness and grammar: The semantics of relative and absolute gradable adjectives." *Linguistics and Philosophy* 30: 1.

Kertscher, T., Friday, March 14, 2014. In context: Were Paul Ryan's poverty comments a 'thinly veiled racial attack'? *Politifact Wisconsin*. Accessed January 26, 2019.
<https://www.politifact.com/wisconsin/article/2014/mar/14/context-paul-ryans-poverty-comment-s-racial-attack/>

Khan, L.A., 2005. The essentialist terrorist. *Washburn LJ*, 45.

Khemlani, S., Leslie, S.J., and Glucksberg, S., 2009. Generics, prevalence, and default inferences. *Proceedings of the 31st Annual Cognitive Science Society*. Amsterdam: Cognitive Science Society.

Khoo, J., 2017. Code words in political discourse. *Philosophical Topics*, 45(2), pp.33-64.

Khosravinik, M., 2010. The representation of refugees, asylum seekers and immigrants in British newspapers: A critical discourse analysis. *Journal of Language and Politics*, 9(1), pp.1-28.

Kim, C.J., 1999. The racial triangulation of Asian Americans. *Politics & Society*, 27(1), pp.105-138.

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

- King, J., 2014. The metasemantics of contextual sensitivity, in Burgess and Sherman (eds.) *Metasemantics: New essays on the foundations of meaning*, pp.97-118.
- King, J. and Stanley, J., 2005. Semantics, pragmatics, and the role of semantic content. *Semantics versus pragmatics*, pp.111-164.
- Kratzer, A., 1995. Individual-level predicates. In *The Generic Book*, edited by Greg Carlson and Francis J. Pelletier, 1–124. Chicago, IL: University of Chicago Press.
- Krifka, M., Pelletier, F.J., Carlson, G., ter Meulen, A., Link, G., and Chierchia, G., 1995. Genericity: an introduction. In *The Generic Book*, edited by G. Carlson and F.J. Pelletier, 1–124. Chicago, IL: University of Chicago Press.
- Kripke, S., 1972. Naming and necessity. In *Semantics of natural language* (pp. 253-355). Springer, Dordrecht.
- Langan, P.A., 1991. America's soaring prison population. *Science*, 251(5001), pp.1568-1573.
- Lasnik, P., 2005. Context dependence, disagreement, and predicates of personal taste. *Linguistics and philosophy*, 28(6), pp.643-686.
- Lazaridou-Chatzigeorgiou, D., Stockall, L. and Katsos, N., 2017. A new look at the “generic overgeneralisation” effect. *Inquiry*, pp.1-27.
- Lebron, C.J., 2013. *The Color of Our Shame: Race and Justice in Our Time*. Oxford University Press.
- Lee, T.L. and Fiske, S.T., 2006. Not an outgroup, not yet an ingroup: Immigrants in the stereotype content model. *International Journal of Intercultural Relations*, 30(6), pp.751-768.
- Leslie, S. J., 2007. Generics and the structure of the mind. *Philosophical Perspectives* 21 (1): 375–403.
- Leslie, S. J., 2008. Generics: Cognition and acquisition. *Philosophical Review* 117 (1): 1–47.
- Leslie, S. J., 2014. Carving up the social world with generics. *Oxford Studies in Experimental Philosophy* 1: 208–231.
- Leslie, S.J., 2015. “Hilary Clinton is the only man in the Obama administration”: Dual character concepts, generics, and gender.” *Analytic Philosophy* Vol. 56, no. 2 pp. 111-141.

Leslie, S. J. 2017. The original sin of cognition: Fear, prejudice, and generalization. *Journal of Philosophy*.

Leslie, S. J., and S. A. Gelman. 2012. Quantified statements are recalled as generics: Evidence from preschool children and adults. *Cognitive Psychology* 64: 186–214.

Levs, J., April 29, 2015. After Baltimore riots, some leaders slam “thug” as the new n-word. *CNN.com*. Accessed January 26, 2019.

<https://www.cnn.com/2015/04/29/us/baltimore-riots-thug-n-word/index.html>

Lewis, B., 1990. The roots of Muslim rage. *The Atlantic Monthly*, 266(3), pp.47-60.

Lewis, D., 1975. Adverbs of Quantification. In *Formal Semantics of Natural Language*, edited by E. L. Keenan, 178–188. Cambridge University Press.

Leyens, J.P., Rodriguez-Perez, A., Rodriguez-Torres, R., Gaunt, R., Paladino, M.P., Vaes, J. and Demoulin, S., 2001. Psychological essentialism and the differential attribution of uniquely human emotions to ingroups and outgroups. *European Journal of Social Psychology*, 31(4), pp.395-411.

Leyens, J.P., Cortes, B., Demoulin, S., Dovidio, J.F., Fiske, S.T., Gaunt, R., Paladino, M.P., Rodriguez-Perez, A., Rodriguez-Torres, R. and Vaes, J., 2003. Emotional prejudice, essentialism, and nationalism The 2002 Tajfel Lecture. *European Journal of Social Psychology*, 33(6), pp.703-717.

Lind, D., Trump on deported immigrants: “They’re not people. They’re animals.” *Vox*. May 17, 2018

Ludlow, P., 2014. *Living words: Meaning underdetermination and the dynamic lexicon*. OUP Oxford.

Maass, A., Salvi, D., Arcuri, L. and Semin, G.R., 1989. Language use in intergroup contexts: The linguistic intergroup bias. *Journal of personality and social psychology*, 57(6), p.981.

MacFarlane, J., 2014. *Assessment Sensitivity: Relative Truth and its Applications*. OUP Oxford.

Mahalingam, R., 2003. Essentialism, culture, and power: Representations of social class. *Journal of Social Issues*, 59(4), pp.733-749.

Maitra, I. and McGowan, M.K. eds., 2012. *Speech and Harm: Controversies Over Free Speech*. Oxford University Press on Demand.

Mallon, R., 2016. *The Construction of Human Kinds*. Oxford University Press.

Mandalaywala, T.M., Amodio, D.M. and Rhodes, M., 2018. Essentialism promotes racial prejudice by increasing endorsement of social hierarchies. *Social Psychological and Personality Science*, 9(4), pp.461-469.

Manne, K., 2017. *Down Girl: The Logic of Misogyny*. Oxford University Press.

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.)

McConnell-Ginet, S., 2003. What's in a name? Social labeling and gender practices, in *The Handbook of Language and Gender*, edited J. Holmes and M. Meyerhoff, 69–97. Oxford: Blackwell.

McConnell-Ginet, S., 2012. Generic predicates and interest-relativity. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, 57(2), pp.261-287.

McGowan, M.K., 2012. On “whites only” signs and racist hate speech: Verbal acts of racial discrimination. *Speech and harm: Controversies over free speech*, pp.121-147.

Mendelberg, T., 2001. *The Race Card*. Princeton: Princeton University Press.

Mercier, B., Norris, A. and Shariff, A.F., 2018. Muslim mass shooters are perceived as less mentally ill and more motivated by religion. *Psychology of Violence*.

Metzl, J.M., October 6, 2017. When the shooter is white. *The Washington Post*. Accessed January 26, 2019.

https://www.washingtonpost.com/news/made-by-history/wp/2017/10/06/when-the-shooter-is-white/?noredirect=on&utm_term=.00ddd81abb60

Milgram, S., 1963. Behavioral study of obedience. *The Journal of abnormal and social psychology*, 67(4), p.371.

Mills, C., 1997. *The Racial Contract*. Ithaca: Cornell University Press.

Mills, C., 1998. *Blackness Visible: Essays on Philosophy and Race*. Ithaca: Cornell University Press.

Monahan, J., and Hood, G.L., 1976. Psychologically disordered and criminal offenders: Perceptions of their volition and responsibility. *Correctional Psychologist*, 3(2), pp.123-134.

Monterosso, J., Royzman, E.B. and Schwartz, B., 2005. Explaining away responsibility: Effects of scientific explanation on perceived culpability. *Ethics & Behavior*, 15(2), pp.139-158.

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

Muhammad, K.G., 2011. *The Condemnation of Blackness*. Harvard University Press.

Muhlebach, D., 2019. Semantic contestations and the meaning of politically significant terms. *Inquiry*.

Newheiser, A.K., Tausch, N., Dovidio, J.F. and Hewstone, M., 2009. Entitativity and prejudice: Examining their relationship and the moderating effect of attitude certainty. *Journal of Experimental Social Psychology*, 45(4), pp.920-926.

Nickel, B., 2008. Generics and the ways of normality. *Linguistics and Philosophy*, 31(6), pp.629-648.

Nickel, B., 2016. *Between Logic and the World: An Integrated Theory of Generics*. Oxford University Press.

Nickel, B. "The Division of Ideological Labor," unpublished manuscript.

Nisbett, R. E. and Ross, L., 1980. *Human Inference: Strategies and Shortcomings of Social Judgment*. (Englewood Cliffs, NJ: Prentice-Hall).

Nosek, B. A., Banaji, M.R., and Greenwald., A.G., 2002. Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice* 6 (1): 101–115.

Nunberg, G., 2005. "When Words Break Down." Accessed January 26, 2019.
<http://people.ischool.berkeley.edu/~nunberg/looting.html>

O'Donnell, P., 2017. Generics, race, and social perspectives. *Inquiry*, pp.1-38.

Office of the Commissioner of Baseball., 2015. Official baseball rules. New York: Major League Baseball Publications.

Omi, M., and Winant, H., 2013. *Racial Formation in the United States*. New York: Routledge.

Painter, N.I., 2010. *The History of White People*. WW Norton & Company.

Peffley, M., Hurwitz, J. and Sniderman, P.M., 1997. Racial stereotypes and whites' political views of blacks in the context of welfare and crime. *American Journal of Political Science*, pp.30-60.

Perlstein, R., Nov. 13, 2012. Exclusive: Lee Atwater's infamous 1981 interview on the Southern Strategy. *The Nation*. Accessed January 26, 2019.
<https://www.thenation.com/article/exclusive-lee-atwaters-infamous-1981-interview-southern-strategy/>

Perry, J., 2001. *Reference and Reflexivity*. Stanford: CSLI.

Pettigrew, T.F., 1979. The ultimate attribution error: Extending Allport's cognitive analysis of prejudice. *Personality and social psychology bulletin*, 5(4), pp.461-476.

Plunkett, D. and Sundell, T., 2013. "Disagreement and the Semantics of Normative and Evaluative Terms." *Philosophers' Imprint* 13.

Prasada, S., and Dillingham, E.M., 2006. Principled and statistical connections in common sense conception." *Cognition* 99 (1): 73–112.

Prasada, S. and Dillingham, E.M., 2009. Representation of principled connections: A window onto the formal aspect of common sense conception. *Cognitive Science*, 33(3), pp.401-448.

Prasada, S., Khemlani, S., Leslie, S.J. and Glucksberg, S., 2013. Conceptual distinctions amongst generics. *Cognition*, 126(3), pp.405-422.

Prentice, D.A. and Miller, D.T., 2006. Essentializing differences between women and men. *Psychological Science*, 17(2), pp.129-135.

Price, Huw. 2007. Causal perspectivalism. In *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*, edited by Huw Price and Richard Corry, 250–292. Oxford University Press.

Price, J. and Wolfers, J., 2010. Racial discrimination among NBA referees. *The Quarterly Journal of Economics*, 125(4), pp.1859-1887.

Prinz, J., 2002. *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge: MIT Press.

Purvis, T. and Hunt, A., 1993. Discourse, ideology, discourse, ideology, discourse, ideology... *British Journal of Sociology*, pp.473-499.

Ramsey, E.M., Achter, P.J. and Condit, C.M., 2001. Genetics, race, and crime: An audience study exploring The Bell Curve and book reviews. *Critical Studies in Media Communication*, 18(1), pp.1-22.

Rawls, J. 1955. Two concepts of rules. *The Philosophical Review* 64 (1): 3–32.

Recanati, F., 2004. *Literal Meaning*. Cambridge University Press.

Rhodes, M. and Gelman, S.A., 2009. A developmental examination of the conceptual structure of animal, artifact, and human social categories across two cultural contexts. *Cognitive psychology*, 59(3), pp.244-274.

Rhodes, M., Leslie, S.J., and Tworek, C., 2012. Cultural transmission of social essentialism. *Proceedings of the National Academy of Sciences* 109 (34): 13526–13531.

Rhodes, M. and Mandalaywala, T.M., 2017. The development and developmental consequences of social essentialism. *Wiley Interdisciplinary Reviews: Cognitive Science*, 8(4), p. 1437.

Rudinger, R., May, C. and Van Durme, B., 2017. Social bias in elicited natural language inferences. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing* (pp. 74-79).

Sagar, H.A. and Schofield, J.W., 1980. Racial and behavioral cues in black and white children's perceptions of ambiguously aggressive acts. *Journal of personality and social psychology*, 39(4), p.590.

Saletan, W., 2017. What Trump supporters really believe. *Slate*, August, 29. Accessed January 26, 2019: <https://slate.com/news-and-politics/2017/08/trumps-bigoted-base-by-the-numbers.html>

Santa Ana, O. 1999. “Like an animal I was treated”: Anti-immigrant metaphor in US public discourse. *Discourse & society*, 10(2), pp.191-224.

Saul, J.M., 2017. Racial figleaves, the shifting boundaries of the permissible, and the rise of Donald Trump. *Philosophical Topics*, 45(2), pp.97-116.

Saul, J., 2017. Are generics especially pernicious?. *Inquiry*, pp.1-18.

Saul, J., 2018. Protean dogwhistles and political manipulation. *Media Ethics, Free Speech, and the Requirements of Democracy*.

Saul, J., 2018. Dogwhistles, political manipulation, and philosophy of Language. *New Works on Speech Acts*, pp.360-383.

Schwartz, I., 2017. Bannon declares war on GOP establishment: “We’re coming after all of them and we’re going to win”. *RealClearPolitics.com*. October 17. Accessed January 27, 2019: https://www.realclearpolitics.com/video/2017/10/10/bannon_declares_war_on_gop_establishment_were_coming_after_all_of_them_and_were_going_to_win.html

Searle, J., 1995. *The Construction of Social Reality*. New York: Free Press.

Sellars, W., 1956. Empiricism and the Philosophy of Mind. *Minnesota studies in the philosophy of science*, 1(19), pp.253-329.

Shelby, T., 2003, June. Ideology, racism, and critical social theory. In *The Philosophical Forum* (Vol. 34, No. 2, pp. 153-188). Oxford, UK and Malden, USA: Blackwell Publishing Ltd.

Sheldon, J.P., Jayaratne, T.E. and Petty, E.M., 2007. White Americans’ genetic explanations for a perceived race difference in athleticism: The relation to prejudice toward and stereotyping of Blacks. *Athletic Insight: The Online Journal of Sport Psychology*, 9(3), pp.31-56.

Sides, J. and Gross, K., 2013. Stereotypes of Muslims and Support for the War on Terror. *The Journal of Politics*, 75(3), pp.583-598.

Silk, A., 2016. *Discourse Contextualism: A framework for contextualist semantics and pragmatics*. Oxford University Press.

Silva, P., 2018. A Bayesian explanation of the irrationality of sexist and racist beliefs involving generic content. *Synthese*, pp.1-23.

Small, D.A., Lerner, J.S. and Fischhoff, B., 2006. Emotion priming and attributions for terrorism: Americans' reactions in a national field experiment. *Political Psychology*, 27(2), pp.289-298.

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

- Sommers, S.R., Apfelbaum, E.P., Dukes, K.N., Toosi, N. and Wang, E.J., 2006. Race and media coverage of Hurricane Katrina: Analysis, implications, and future research questions. *Analyses of Social Issues and Public Policy*, 6(1), pp.39-55.
- Sperber, D., 1994. The modularity of thought and the epidemiology of representations. *Mapping the mind: Domain specificity in cognition and culture*, pp.39-67.
- Stalnaker, R. 1978. Assertion. In P. Cole (ed.) *Syntax and Semantics*, Vol. 9. New York: Academic Press.
- Stalnaker, R. 1997. Reference and Necessity. In B. Hale and C. Wright (eds.), *The Blackwell Companion to the Philosophy of Language*. Oxford: Blackwell, pp. 534-54.
- Stalnaker, R., 2002. Common ground. *Linguistics and philosophy*, 25(5-6), pp.701-721.
- Stanley, J., 2000. Context and logical form. *Linguistics and philosophy*, 23(4), pp.391-434.
- Stanley, J., 2015. *How propaganda works*. Princeton University Press.
- Stanley, J. and Gendler Szabó, Z., 2000. On quantifier domain restriction. *Mind & Language*, 15(2-3), pp.219-261.
- Stephenson, T., 2007. Judge dependence, epistemic modals, and predicates of personal taste. *Linguistics and Philosophy*, 30(4), pp.487-525.
- Sterken, R.K., 2015a. Leslie on generics. *Philosophical Studies*, 172(9), pp.2493-2512.
- Sterken, R. K., 2015b. Generics in Context. *Philosopher's Imprint*. Vol. 15, No. 21.
- Swanson, E., 2018. Slurs and ideology. In Celikates, R., Haslanger, S., and Stanley, J., (eds.), *Ideology*.
- Szabó, Z.G., 2001. Adjectives in context. *Pragmatics and Beyond New Series*, pp.119-146.
- Taylor, D.M. and Jaggi, V., 1974. Ethnocentrism and causal attribution in a South Indian context. *Journal of cross-cultural Psychology*, 5(2), pp.162-171.
- Taylor, J., 2018. Trump tests midterm message on immigration, MS-13 “animals” during Tenn. rally. *NPR.org*, May 29. Accessed January 27, 2019:
<https://www.npr.org/2018/05/29/615355282/trump-tests-midterm-message-on-immigration-ms-13-animals-during-tenn-rally>

- Taylor, K., 2007. A little sensitivity goes a long way. *Context-sensitivity and semantic minimalism: New essays on semantics and pragmatics*, pp.63-93.
- Taylor, P.C., 2004. *Race: A Philosophical Introduction*. Polity.
- Teichman, M., 2017. The sophisticated kind theory. *Inquiry*, pp.1-47.
- Thomas, J. and Boehlefeld, S., 1991. Rethinking abolitionism: What do we do with Henry? Review of de Haan, *The Politics of Redress*
- Travis, C., 1985. On what is strictly speaking true. *Canadian Journal of Philosophy*, 15(2), pp.187-229.
- Tversky, A. and Kahneman, D., 1983. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological review*, 90(4), p.293.
- U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics. 2011. Homicide trends in the United States, 1980–2008. Accessed January 27, 2019: <http://www.bjs.gov/content/pub/pdf/htus8008.pdf>
- Van Dijk, T.A., 2000. New(s) racism: A discourse analytical approach. *Ethnic minorities and the media*, 37, pp.33-49.
- Veltman, F., 1996. Defaults in update semantics. *Journal of philosophical logic*, 25(3), pp.221-261.
- Von Fintel, K. and Gillies, A.S., 2011. Might made right. *Epistemic modality*, pp.108-130.
- Wacquant, L., 2009. *Punishing the Poor: The Neoliberal Government of Social Insecurity*. Duke University Press.
- Wagner, K., 2014. The word “thug”: Regressing. *Deadspin.com*. Accessed January 21. <http://regressing.deadspin.com/the-word-thug-was-uttered-625-times-on-tv-yesterday-1506098319>.
- Wason, P. C. 1968. Reasoning about a rule. *Quarterly Journal of Experimental Psychology*. **20** (3): 273–281.
- Watson, B., Jones, D., and Robinson-Saunders, P., 1995. Drug use and African Americans: Myth versus reality. *Journal of Alcohol and Drug Education* 40 (2): 19–39.

- Waxman, S. R., 2010. Names will never hurt me? Naming and the development of racial and gender categories in preschool-aged children. *European Journal of Social Psychology*, 40(4), pp.593-610.
- Weaver, V.M., 2007. Frontlash: Race and the development of punitive crime policy. *Studies in American political development*, 21(2), pp.230-265.
- Weiner, B., Perry, R.P. and Magnusson, J., 1988. An attributional analysis of reactions to stigmas. *Journal of personality and social psychology*, 55(5), p.738.
- White, I.K., 2007. When race matters and when it doesn't: Racial group differences in response to racial cues. *American Political Science Review*, 101(2), pp.339-354.
- Williams, M.J. and Eberhardt, J.L., 2008. Biological conceptions of race and the motivation to cross racial boundaries. *Journal of personality and social psychology*, 94(6), p.1033.
- Wilson, M., 1982. Predicate meets property. *The Philosophical Review*, 91(4), pp.549-589.
- Wilson, Ryan. January 22, 2014. Sherman: "Thug" is accepted way of calling someone n-word." *CBSsports.com*. Accessed January 26, 2019.
<https://www.cbssports.com/nfl/news/richard-sherman-thug-is-accepted-way-of-calling-someone-n-word/>
- Wodak, R., 2015. *The politics of fear: What right-wing populist discourses mean*. Sage.
- Wodak, Ruth. 2017. The "establishment," the "elites," and "the people": Who's who. *Journal of Language and Politics*.
- Yin, Tung. 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.
- Yzerbyt, V., Rocher, S., & Schadron, G. 1997. Stereotypes as explanations: A subjective essentialistic view of group perception. In R. Spears, P. J. Oakes, N. Ellemers, & S. A. Haslam (Eds.), *The social psychology of stereotyping and group life* (pp. 20-50). Malden, : Blackwell Publishing.
- Yzerbyt, V.Y., Rogier, A. and Fiske, S.T., 1998. Group entitativity and social attribution: On translating situational constraints into stereotypes. *Personality and Social Psychology Bulletin*, 24(10), pp.1089-1103.

Yzerbyt, V., Corneille, O. and Estrada, C., 2001. The interplay of subjective essentialism and entitativity in the formation of stereotypes. *Personality and Social Psychology Review*, 5(2), pp.141-155.

Zinn, H., 1968. *Disobedience and democracy: Nine fallacies on law and order* (Vol. 4). South End Press.

Curriculum Vitae Patrick O'Donnell

Department of Philosophy
Johns Hopkins University
3400 N. Charles St.
Baltimore, MD 21218

podonn14@jhu.edu
www.patrickaodonnell.com
443-690-7311

AREAS OF SPECIALIZATION

Philosophy of Race, Philosophy of Language

AREAS OF COMPETENCE

Moral Philosophy, Philosophy of Mind, Feminist Philosophy, Metaphysics (esp. Social Ontology), Phenomenology, Philosophy of Technology

EDUCATION

PhD in Philosophy, Johns Hopkins Universityreceived February 14, 2019

Dissertation: *Racialized Discourse at the Intersection of Meaning, Mind, and Metaphysics*

Committee: Steven Gross (director), Christopher Lebron, Justin Bledin, Lester Spence, Bernhard Nickel

Certification, Johns Hopkins Teaching Academy.....2018

MA in Philosophy, Johns Hopkins University.....2013

Third-Year Paper: "The Psychology and Pragmatics of Racial Kind Generics"

MA in Philosophy, *magna cum laude*, Katholieke Universiteit Leuven.....2010

Thesis: "Mental Representation and the Embodied, Embedded Mind"

Advisor: Filip Buckens

BA in Philosophy, minor in music, *cum laude* Columbia University.....2009

Thesis: "Heidegger's Unity of Thought: Worldhood in *Being and Time* and in the Age of Technology"

Advisor: Taylor Carman

PUBLICATIONS (* denotes refereed)

Academic

*"Generics, Race, and Social Perspectives," *Inquiry*. Special issue on genericity in philosophy and linguistics, eds. Rachel Katherine Sterken and Bernhard Nickel. 2017.

Non-Academic

“Still Dreaming: Review of Ta-Nehisi Coates’ *Between the World and Me*,” *America*, vol. 214, no. 9, 32+, March 3, 2016. (1500 words). Available at:

<https://www.americamagazine.org/issue/culture/still-dreaming>

WORK IN PROGRESS

“Not-so-Coded Words” (R+R, under revision)

“White Supremacy as Metaphysics” (in preparation)

“Racialized Terms and Inherent Genericity” (drafting stage)

“Social Construction and Shifty Racialized Categories” (drafting stage)

DISSERTATION ABSTRACT

Racialized Discourse at the Intersection of Meaning, Mind, and Metaphysics

Racialized discourse is language that transmits potentially harmful representations of racial groups. It is also a tool for maintaining status quo racial hierarchies. A theory of racialized discourse should explain what the content of these transmitted representations are, how they are transmitted in communication, and how they manage to sustain racial hierarchies. I meet this explanatory requirement via an original account of the semantics of racially stereotypical generics (e.g. “Blacks are aggressive,” “Muslims are terrorists,” “Immigrants are violent”) and racialized terms (or “code words”) deployed in the context of political discourse (e.g. “thug,” “terrorist,” “inner city,” “welfare.”) These types of language purport to say something about the members of racial groups are like *by nature*. They thus essentialize these groups, and indirectly increase tolerance for social hierarchy.

AWARDS, HONORS, GRANTS

APA Graduate Student Travel Grant for Eastern APA, January 2019

Dean’s Teaching Fellowship Prize (freshman seminar version of DTF class), Fall 2018

David Sachs Dissertation Completion Fellowship, Spring 2018

Dean’s Teaching Fellowship (graduate school-wide teaching fellowship), Fall 2017

Miller Prize (JHU graduate student paper award), 2015

PRESENTATIONS (* denotes refereed)

“Not-so-coded words”

-APA Eastern Division Meeting, New York, January 2019*

-12th Inter-University Workshop: Language and Power, Barcelona, May 2018*

-Princeton Workshop in Social Philosophy, Princeton, February 2018*

“White Supremacy as Metaphysics”

- International Society for Social Ontology, Tufts University, August 2018*
- Hammond Society Graduate Colloquium, Johns Hopkins, April 2016

“Racialized Terms and Inherent Genericity”

- Generic Generalizations: Meaning and Social Practices, University of Sherbrooke, March 2018*
- GRSelona 3, University of Barcelona, June 2016*

“(Pseudo)gradability in the Semantics of Racialized Terms” (poster)

- Society for Philosophy and Psychology, Johns Hopkins, June 2017*

“Frantz Fanon’s *Black Skin, White Masks*” (guest lecture)

- Course: Philosophy of Oppression and Resistance, Johns Hopkins, July 2016

“Thugs: Social Negotiation and Linguistic Resistance”

- 3rd Philosopher’s Cocoon, University of Tampa, November 2015*

“Behaviorism: Philosophical Hurdles and the Skinner-Chomsky Debate” (guest lecture)

- Course: History of Psychology, Johns Hopkins, March 2015

“Ideology, Racialization, and Semantics”

- Race and Racism in Comparative Perspective, Johns Hopkins, April 2016*
- Prometheus Philosophy Colloquium, Johns Hopkins, October 2015
- Hammond Society Graduate Colloquium, Johns Hopkins, April 2015

“Social Kind Generics”

- Prometheus Philosophy Colloquium, Johns Hopkins, April 2014

“Artifacts at Auschwitz: Trauma, Worldhood, and the Ready-to-Hand”

- Phenomenology and Religion Conference, Katholieke Universiteit Leuven, May 2010

COMMENTS AND PANELS

Comments on Brian Donohue’s “Eliminating the Descent Condition from Gracia’s Genetic Common Bundle View of Race,” Eastern American Philosophical Association, January 2017

Comments on Matthew Schlesinger's "Poetics and Praxis: Césaire and Fanon on Decolonizing Consciousness," Fourth Mid-Atlantic Undergraduate Philosophy Conference, Johns Hopkins, April 2016

Comments on Patrick Kenney's "Income Inequality as a Restriction of Freedom," Fourth Mid-Atlantic Undergraduate Philosophy Conference, Johns Hopkins, April 2016

Panel Discussant, "Fighting Back: Responses to State Failure, Violence, and Political Discrimination," Race and Racism in Comparative Perspective, Johns Hopkins, December 2015

Comments on Dax Bennington's "Cognitive Penetration and Phenomenal Conservatism: A Reply to McGrath," Philosopher's Cocoon Conference, November 2015

Comments on Daniel Estrada's "Autonomy in Non-Biological Systems," Nature and Norms Grad Conference, Johns Hopkins, October 2015

Comments on Charles Dalrymple-Fraser's "Ableism in the Abortion Debate," Third Mid-Atlantic Undergraduate Philosophy Conference, Johns Hopkins, April 2015

TEACHING

As primary instructor

Johns Hopkins

The Philosophy of Race and Racism (Winter 2015, Fall 2017, Fall 2018)

Philosophy of Oppression and Resistance (Winter 2016)

University of Maryland, Baltimore County (adjunct)

Introduction to Philosophy (Spring 2019)

Feminist Philosophy (Fall 2018)

The Philosophy of Race and Racism (Spring 2017)

Introduction to Moral Theory (Fall 2016, Spring 2017)

Towson University (adjunct)

Introduction to Ethics (Spring 2019)

Johns Hopkins Center for Talented Youth (summer instruction for gifted students, aged 12-17):

Human Nature and Technology (Summer 2015, Summer 2016)

As graduate instructor

Johns Hopkins

Introduction to Formal Logic, with Peter Achinstein (Spring 2017)
Introduction to Bioethics, with Hilary Bok (Fall 2013, Fall 2016)
Introduction to Moral Philosophy, with Nandi Theunissen (Spring 2014, Spring 2016)
Introduction to Philosophy of Mind, with Meredith Williams (Fall 2015)
History of Modern Psychology, with Paul Hofer (Fall 2015)
Introduction to Formal Logic, with Justin Bledin (Spring 2015)
Philosophical Problems, with Steven Gross (Fall 2014)

Johns Hopkins Center for Talented Youth (instruction for gifted students, aged 12-17)

Introduction to Logic, with Daniel Hicks (Summer 2013)
Philosophy of Mind, with Daniel Estrada (Summer 2012)
Human Nature and Technology, with Daniel Estrada (Summer 2011, Summer 2012)
Introduction to Philosophy, with Bernie Cantens (Summer 2011)

As tutor

Symbolic Logic, Arete Educational Counseling, NY (2012-2013)

PROFESSIONAL SERVICE

Referee: *Synthese*

Conference Organizer: “Nature and Norms: Values in a Material World,” Grad Conference, Johns Hopkins University, October 2015

Conference Organizer: “Phenomenology and Religion,” Grad Conference, Katholieke Universiteit Leuven, May 2010

PROFESSIONAL MEMBERSHIPS

American Philosophical Association
International Society for Social Ontology

REFERENCES

Steven Gross

Professor of Philosophy
Johns Hopkins University
sgross11[AT]jhu.edu

Steve Yalowitz (teaching reference)

Professor of Philosophy
University of Maryland, Baltimore County
yalowitz[AT]umbc.edu

Nandi Theunissen

Luvell Anderson

Associate Professor of Philosophy
University of Pittsburgh
lit44[AT]pitt.edu

Justin Bledin
Assistant Professor of Philosophy
Johns Hopkins University
jbledin[AT]jhu.edu

Assistant Professor of Philosophy
Syracuse University
lander04[AT]syr.edu

Daniel Estrada (teaching reference)
Faculty Lecturer
New Jersey Institute of Technology
djestrada@gmail.com