University of Wisconsin Milwaukee

UWM Digital Commons

Theses and Dissertations

May 2020

How Code Words Work

Farhad Taraz
University of Wisconsin-Milwaukee

Recommended Citation

HOW CODE WORDS WORK

by

Farhad Taraz

A Thesis Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Master of Arts

in Philosophy

at

The University of Wisconsin-Milwaukee

May 2020

ABSTRACT

HOW CODE WORDS WORK

by
Farhad Taraz

The University of Wisconsin-Milwaukee, 2020
Under the Supervision of Professor Peter van Elswyk

This paper aims at giving a novel theory of code words. I start with an explanation of what code words are by introducing four primary features of them. Then, I turn to three major theories of code words: those offered by Stanley, Khoo, and Saul. I show how each of these theories falls short of giving a thorough account of all features of code words. Finally, drawing on these lessons, I will formulate a new theory which suggests that code words are polysemous.

To those who do not have a voice to protect themselves from the harms of malicious

speeches and malevolent acts.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I am genuinely grateful for the chance to spend two wonderful years in the philosophy department at UWM, where I not only learned how to practice philosophy but also how to be a better human (although moral knowledge does not bring moral action). Though my life in Milwaukee was not always warm and bright like a breezy summer day by Bradford Beach, it also was not always gloomy and dark as a biting winter night on Downer Avenue. As much as those bright days reinforced my deep commitment to working hard and being optimistic, those gloomy days taught me to be aware of the fragility of human existence and the perilous nature of things that I used to take for granted. It was an uphill struggle, but it was all worth it.

Fortunately, I started my studies at UWM when Peter van Elswyk began his career there. The man whose vigorous lectures and constant pop-culture references initially intimidated me ended up being my mentor and advisor. I cannot be grateful enough for his boundless support, both when he was encouraging me in things as small as obeying obscure grammar rules, and when he was repeatedly challenging me to not sacrifice clarity for an ambition of generality. His neat office was where the idea for this project emerged in Fall 2018, and it was where- besides occasional cafes- I took advantage of his helpful comments throughout this project. For all this, and for the fact that he believed in me, I will always be in debt to him.

Also, I was lucky that in the first semester I took a course with Joshua Spencer on the metaphysics of vagueness. Even though the topic made me distance myself more steadfastly from metaphysics, Joshua's exceptional clarity and wonderful character manifested the Platonic Idea of a philosophy teacher for me. I will never forget how terrible

INTRODUCTION

Words can harm. Using code words in political discourse is one of the many ways by which verbal communication impacts certain groups of society in a negative way. In this paper, I will focus on racial code words as one important instance of code words and in the rest of the paper by 'code word' I mean racial code word. However, I think the results of this paper are true for other kinds of code word as well. Drawing on empirical studies on code words, I begin my discussion by giving a picture of what code words are. I will introduce four primary features of code words that cry out for explanation. Then, in section two, I will look at the main theories that are offered to explain the mechanism of code words. Stanley's[1] semantic theory of code words, and Khoo[2] and Saul's[3] pragmatic theories are subjects of my discussion. My aim is to show how each of these theories falls short of a thorough explanation of the features of code words. Drawing on these lessons, in section three, I will formulate a new theory which suggests that code words are polysemous.

---

[1]Stanley 2015.
[2]Khoo 2017.
[3]Saul 2018.

CHAPTER 1

CODE WORDS: WHAT THEY ARE

In this section, I am going to illustrate code words by isolating four primary features of them. Let's start our discussion with an example of the use of a racial code word in a political speech delivered by former Representative Paul Ryan:

> 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.[1]

This speech, and ones like it, have four features that a theory of code words needs to explain. I should note that while there is not a clear definition of code words in the literature, I provide this list of features that, I believe, covers the most substantial aspects of code words.

The first and most noticeable aspect of code words is that they cause a racially negative effect. Using code words under certain conditions in a speech can bring about attitudes that are biased against a certain racial identity. Hurwitz and Peffley have shown this effect of code words by:

> embedding an experiment in a national survey of whites, where a random half

---

[1] Delievered in a radio talk show hosted by Bill Bennett on March 12, 2014.

of respondents was asked whether they support spending money for prisons (versus antipoverty programs) to lock up 'violent criminals,' while the other half was asked about 'violent inner city criminals.' [.....] [They found] that whites' racial attitudes (for example, racial stereotypes) were much more important in shaping preferences for punitive policies when they receive the racially coded, 'inner city' question.[2]

Psychologists explain what is happening here not in terms of new beliefs that people acquire in the process of hearing the coded speech, but instead they interpret it in terms of the activation of a psychological mechanism called 'racial resentment'.[3] Tesler and Seans describe racial resentment in terms of the following four claims: "(1) blacks no longer face much discrimination, (2) their disadvantage mainly reflects their poor work ethic, (3) they are demanding too much too fast, (4) they have gotten more than they deserve."[4] They show in their research that white Americans, overall, get high scores when their racial resentment is tested. A crucial point about racial resentment is that although people might not hold any explicit racist beliefs, they have this racial resentment to some extent.[5] What code words do is evoke this racial resentment.

In our example, the problematic effect that is produced by Mr. Ryan can be characterized as the following message: "black people are lazy and do not want to work" which is the second component of racial resentment. This effect was detected by Representative Barbara Lee as a "thinly veiled racial attack". She said: "when Mr. Ryan says 'inner city', when he says 'culture', these are simply code words for what he really means: 'black.'" In sum, the use of a code word can be followed by an effect that is harmful for certain racial groups. *Call this the negative racial effect,* or the racial effect for short.[6]

---

[2] Horwitz and Peffley 2005, p. 99.

[3] Mendelberg 2001. p.7

[4] Tesler and Sears 2010. p. 18.

[5] *ibid*, p. 19.

[6] When I say the racial effect is produced by code words, I do not mean such an effect cannot be seen

However, the racial effect does not always come with using those words. There are at least two types of cases of using code words in which the effect is not present. The first type of case is when it becomes explicit that the used code word aims at bringing about the racial effect. As Saul puts it, "[t]he priming of racial resentment only works if it remains covert."[7] In this type of case, the coded message functions like the explicit racial message and does not prime racial resentment in a distinct way anymore.[8]

The second type of case occurs when the context is not prone to trigger racial biases by virtue of code words. Consider these two contexts, for instance. First, we can think of the graduation party of a black woman in which her white and black friends are present. When a speaker who gushes with praise about the woman who used to live in the inner city and now has achieved academic success, it is hard to imagine that this speech would prime any racial resentment in her white friends' minds. Second, we can also think of two white city planners who are talking about the most efficient route for city's new trolley service. In this context, talking about inner city does not mean anything other than one part of the city and will not prime racial resentment.

So, we know that there are cases in which code words are used but the racial effect does not follow, either because it has become explicit by discussing it or because it is used in a context that is not prone to trigger racially negative attitudes. *Call this the non-necessitation condition.*

---

at all in the absence of those words. What is special about these words is that when they are used, racial resentment becomes significantly higher than normal. So, in this paper when I am talking about the presence or absence of the racial effect, I do not mean mere existence or absence of the racial resentment, what I mean, instead, is significantly higher rate of racial resentment.

[7] Saul 2018, p.374.

[8] This effect was first seen after the airing of the Willie Horton ad. This ad was aired during the 1988 Bush campaign. Willie Horton was a felon convicted a life sentence for murder. While he was a beneficiary of Massachusetts weekend furlough program, he committed assault and robbery and rape. Since Michael Dukakis, who was running for presidential election against Bush, was the governor of Massachusset at the time of Horton's release and he supported furlough program, this ad was claimed to be an ad against him on the basis of him being easy on criminals. However, showing a black criminal person was highly effective on fueling racially biased voters. However, when Jesse Jackson started to criticize it on racial grounds, the ad was not effective anymore.

Another important aspect of code words is that while they cause the racial effect, they also provide room for the speaker to deny that they meant to cause any racial effect or convey any racial meaning. In our example, when Ryan was criticized by Lee, he responded that by saying: "She knows that I don't have a racist bone in my body"; further, he'd like to "get beyond throwing baseless charges at people."[9] Regardless of the real intention of Mr. Ryan, an interesting fact about code words is that they make it feasible to have such debates about them. Code words have this potentiality to be deployed by speakers who calculate that the racial effect would be to their benefit, however, they know that if they are accused to be racist, it would have an unpalatable effect on their reputation. Code words help them to have the good of both worlds, gaining the favorable racially biased result while being able to deny being racist. *Call this the deniability condition.*

Another feature of code words is that the racial effect is promulgated by code words, and neither by other coextensive words nor by words that explicitly refer to a racial group. That is, if the speaker explicitly uses racial words, or uses a word that refers to the same population that the term 'inner city' does, they would not get the same result.[10] This feature of code words is shown in a study by White. Here are the findings of his research:

> I test the effects of explicitly racial [i.e. 'African American'], implicitly racial,
> [i.e. 'inner city'], and non-racial [i.e. 'American' and 'poor'] verbal cues on both
> Black and White Americans' assessments of an ostensibly non-racial issue. The
> results point to important racial differences in the effectiveness of explicit and
> implicit racial verbal cues in activating racial thinking about an issue. Only
> frames that provide oblique references to race [i.e. code words] successfully
> activated racial out-group resentment for Whites.[11]

---

[9]In an interview with Bill O'Reilly on FoxNews: https://www.usatoday.com/story/news/politics/onpolitics/2014/03/26/paul-ryan-barbara-lee-poverty-race/81579832

[10]Again, the claim is not that by using other coextensive words or explicit racial words racial resentment would disappear. What is special about code words is that using them significantly increases racial resentment.

[11]White 2007. p. 339

Drawing on this research, the last feature of code words that needs to be explained is that the racial effect cannot be generated by other words that pick the same population in the world or words that explicitly refer to the targeted racial group. *Call this the uniqueness condition.*

To summarize, we now know that code words have four primary features. The first and foremost aspect of code words is that they cause a negative racial effect, due to the priming of racial resentment of the audience. Second, this effect does not always follow by any use of code words, it is not necessitated when it becomes explicit or when it is used in some contexts. Third, when the effect is produced, it is always possible for the speaker to deny any intention related to racial issues, i.e. they are deniable. Fourth, code words are distinctive from other coextensive words or words that explicitly refer to a race in causing the racial effect, they are unique in this sense.

CHAPTER 2

THEORIES OF CODE WORDS

As a result of our discussion thus far, we know that a theory of code words needs to explain four conditions, racial effect, non-necessitation, deniability, and uniqueness. Now we turn to three major theories that are advanced in order to explain code words. The first account discussed here is a semantic account proposed by Stanley. His theory recognizes the negative effect of code words as a result of a special kind of semantic content that those words have. The other two theories, on the other hand, are pragmatic. They conceive of the negative effect as something that is *done* by code words and has nothing to do with their semantic content. The first theory of this kind, offered by Khoo, suggests that code words give rise to certain racial inferences, given the existence of certain beliefs about those words. The second theory of this kind, put forth by Saul, indicates that code words unconsciously make preexisting racial attitudes salient. In what follows, I will present these theories respectively and try to show how each of them falls short of giving a full account of code words.

2.1   Stanley: Conventional Not-at-Issue Content

In his recent book, *How Propaganda Works*[1], Stanley discusses code words as one important mechanism by which propaganda functions and gives an account of the linguistic process of code words. In order to explain it, Stanley uses the concept of common ground of a conversation. According to Stalnaker, common ground is "the mutually recognized shared information in a situation in which an act of trying to communicate takes place."[2] As the conversation proceeds, according to this picture, the content of a successful assertion "will become the part of the body of information that provides the context for the subsequent discourse."[3] An expression containing code words is also an assertion that is meant to be added to the common ground. Stanley observes that coded speeches have two aspects: one that erodes respect for certain groups of society, and the other is a simple contribution to the debate at issue.[4] The former effect is due to what he calls the negative social meaning of code words. That is, the racial effect, according to Stanley, is caused by this negative social meaning of code words that gets added to the common ground.

Furthermore, in order to address the deniability condition, he draws a distinction between the way that the reasonable content of coded assertions gets added to the common ground and the way that the problematic content does. For this end, Stanley helps his theory to the concepts of at-issue and not-at-issue content. "The at-issue content is what is at issue in the debate."[5] To address what is at issue in the debate by an assertion "is to propose to add it to the common ground."[6] The reasonable aspect of the coded speech is its at-issue content. On the other hand, not-at-issue content "is not advanced as a proposal of a content to be added to the common ground. Not-at-issue content is directly added to

---

[1]Stanley 2015
[2]Stalnaker 2002, p. 704
[3]Stalnaker 1996. p. 98
[4]Stanley 2015, p.130
[5]*ibid*, p. 134
[6]*ibid*, p. 134

the common ground."[7] The derogatory content of a coded speech due to the negative social meaning of code words is the not-at-issue content of it. Since not-at-issue content is "not-negotiable, not directly challengeable, and [is] added [to the common ground] even if the at-issue proposition is rejected"[8] it makes it deniable.

In addition, by explaining how the negative social meaning gets attached to code words, Stanley specifies the kind of not-at-issue content that he has in mind and also addresses the uniqueness condition:

> One kind of linguistic propaganda involves repeated association between words and social meanings. Repeated association is also the mechanism by which *conventiona*l meaning is formed.... My claim in this chapter is that when propagandists use repeated association between words and images, they are forming connections that serve as the basis of *conventional* meaning. At some point, the repeated associations are part of the meaning, the not-at-issue content.[9] (Emphasis added)

So, if the derogatory content is part of the conventional not-at-issue meaning of code words, it makes sense that only code words, and not their coextensive words, can cause the racial effect, because conventional meaning is not necessarily shared between coextensive words. This enables the theory to address the uniqueness condition as well.

To recapitulate, had Ms. Lee read Stanley's book, she would have given a description as to what Mr. Ryan has done as follows. By uttering the sentence about lack of the culture of work in inner cities, his sentence adds two different propositions to the common ground. One proposition is about inner cities and is at-issue and questionable. Another proposition that has a derogatory racial content due to the conventional racial

---

[7]*ibid*, p. 134

[8]Murray, Sarah. "Varieties of Update." *Semantics and Pragmatics*, 7, no. 2, 2014, 1– 53. Quoted from Stanley 2015, p. 135

[9]Stanley 2015, p. 138

meaning of the term 'inner city', is not-at-issue, unquestionable, and directly gets added to the common ground.

Here is how Stanley's theory falls short of capturing all features of code words.

Problems regarding the deniability condition: Here are the resources that Stanley uses to address the deniability condition. Derogatory content of code words is their conventional meaning, and conventional meaning is not-at-issue and, therefore, not up for debate and challenge. However, these resources are not enough to fully explain the deniability condition. The problem with code words is that the speaker who uses them is able to vehemently oppose that the word has any racial connotation. In our example, while Ms. Lee claims that 'inner city' simply means 'black', Mr. Ryan would just respond: "No, it does not. 'Inner city' just refers to a poor neighborhood and has nothing to do with race." However, Stanley's theory does not predict such an answer because if the racial content was part of the conventional meaning of 'inner city' it would not be easy for Ryan to deny *any* connection between them.

To illustrate, take the famous example of the conventional meaning of the word 'therefore', offered by Grice[10], in this sentence:

He is an English man. Therefore, he is brave.

In this case, the conventional meaning of the word 'therefore' implies the meaning that "being brave follows from being English." This conventional not-at-issue content is not challengeable in the sense that it does not seem felicitous for one to respond to this sentence by saying: "no! Being brave does not follow from being English." However, when one does respond so, and this is a crucial point about the conventional meaning, the door is not open for the speaker to deny the conventional meaning of 'therefore'. To put it differently, the sense in which conventional not-at-issue content is not negotiable and

---

[10]Grice 1989. p. 25

challengeable is different from the sense in which code words are not challengeable. In the former sense non-challengeability is about the felicity of challenging the not-at-issue content. However, in the latter sense, non-challengeability is regarding the very conventional meaning of the code word.

Problems regarding the non-necessitation condition: The other objection which is raised by both Khoo and Saul is based on the fact that Stanley's theory cannot explain the non-necessitation condition. As Saul puts it, "the effect of [code words] are not quite so monolithically negative as Stanley takes them to be."[11] She says, we can easily imagine cases in which one is using code words, e.g. inner city, and no one is getting a negative racial message from the sentence.

Khoo puts this objection in a more technical way. He argues that non-cancellability is an inherent property of conventional implicatures and if the derogatory content that Stanley attributes to code words were part of their conventional meaning, this content could not be cancellable. Grice introduced the cancellability test.[12] In the case of conventional implicature, the test is supposed to show that conventional meaning of a word is always implicated by saying it. Khoo uses this criterion to test sentences including 'inner city' and claims that the racial content of these sentences is cancellable. Hence, the content is not their conventional not-at-issue content. Here are Khoo's examples:

1. Food stamp programs help many inner city families, most of whom are white.

2. # Food stamp programs help many inner-city families, most of whom are wealthy suburbanites.[13]

Juxtaposing these two sentences, he argues that while cancelling the racial content in the first sentence does not generate contradiction, cancelling the economic content does.

---

[11]Saul 2018, p. 374
[12]Grice 1989, p. 39
[13]Khoo 2017, p. 46

Consequently, although the latter is part of the conventional meaning of the code word 'inner city' there is no racially derogatory content under the conventional meaning of this word.[14]

Problem regarding the racial effect: Another problem that Saul raises for Stanley's account is related to the way that his theory construes the first condition. That is, the negative effect of code words, according to Stanley, is generated by adding racial propositions to the common ground in a hidden way. However, Saul thinks that what code words do is just make the preexisting racial attitudes salient, and nothing new in terms of racial beliefs is added to the common ground by those words.

2.2 Khoo: Inference Theory

Having diagnosed Stanley's account as problematic, Khoo develops another theory of code words. Khoo's alternative proposal for the mechanism of code words is that they give rise to certain inferences with racial content due to the preexisting beliefs of the audience about those words. The general idea of Khoo's theory is that when speakers use code words the audience has preexisting beliefs about the subject matter of the word or about the word itself. These beliefs lead them to infer a further statement which was neither part of the explicit meaning of the asserted statement nor has any direct relation with the meaning of the code word. According to the theory which he calls the "simple theory of code words"[15]:

The take-home message is that code words need not be code for anything to

---

[14]I think the way that Khoo motivates his point is problematic because the second sentence does not show anything about the relation between 'inner city' and 'poor'. The reason is that there are two other sources of contradiction in the sentence, i.e. 'suburbanite' and 'food stamp'. That is, merely talking about 'inner city' and 'suburbanite', on the one hand, and 'food stamp' and 'wealthy', on the other, might generate contradiction. As a result this sentence is not a good example at all to test the relation between 'poor' and 'inner city'.

[15]*ibid*, p. 35

have 'coded' effects. Here is what is happening, schematically. Take some code word 'C'. We distinguish the preexisting belief about Cs from what is explicitly said and what is inferred on that basis:

- Explicit Statement: x is C.

- Existing Belief: If something is C, then it is R.

- Inferred: x is R.[16][17]

So, in our example, when Ryan says "inner city people lack the culture of work", since people have a preexisting belief about inner cities, i.e. "inner city is populated by black people", they then draw the following racial inference: "black people lack the culture of work." As Khoo puts it, "though [the speaker] does not communicate anything about race, since many hearers have preexisting beliefs about the subject matter of what he does say, [his] speech has the result that hearers will infer some racial belief about the policy from what he says." [18]

Like Stanley, Khoo thinks that the racial effect is produced due to a derogatory content. This content is inferred by the audience as a result of the speakers' using a code word. This inference could take place consciously or unconsciously.[19] The advantage of his theory is that it can aptly explain the deniability condition, because it shows how the speaker could resort to the beliefs of the audience as the source of the derogatory content and, hence, the racial effect. That is, if the speaker is accused of being racist, they can claim that the derogatory content is inferred by the accuser.

However, Khoo's simple theory needs more resources to explain the uniqueness

---

[16] *ibid*, p. 50

[17] The schematic way by which Khoo tries to capture the inference that, according to his theory, we draw in the natural language does not seem right. He should have said: 'C is x', and, 'C is R', therefore, 'R is x'. Obviously, this is not a valid argument. However, nobody thinks that all the inferences that the ordinary people tend to draw are based on valid arguments.

[18] *ibid*, p. 48

[19] *ibid*, p. 50

condition. The natural question that comes to mind is: why do people draw racial inferences from a specific word and not its coextensive counterparts? In order to address this worry, Khoo helps himself to the notion of fragmentation. He defines fragmentation as follows: "a rational individual may believe x is F when thinking of x in one way, and believe that x is not F when thinking of x in a different way."[20] For example, it is possible to imagine that people have a racial judgment about a part of the city when it is referred to as inner city, but they do not hold the same racial judgment about the same part of the city when it is referred to as a poor urban area. Khoo believes that his theory supplemented with fragmentation is able to explain the uniqueness of code words.

To summarize his theory, if Lee wanted to give a Khoo-like explanation of what Ryan did, she would have said something as follows. The problem with Mr. Ryan's speech is that he uses the word 'inner city' in his speech, knowing that people make certain racial judgments about that word. As a result, when they hear his speech containing 'inner city' they will draw racial inferences.

Nevertheless, this theory is not able to explain all the features of code words.

Problems regarding the uniqueness condition: The theory falls short of fully addressing the uniqueness condition in two ways. The first problem that uniqueness poses for this theory is that although fragmentation gives the theory a way to explain *how* code words might be unique, there are different ways to fill out the details as to *why* it is so. I do not mean that a theory of code words needs to provide a genealogical story about the process in which code words attain their features. However, "preexisting beliefs about words or their subject matter", is too big of an explanation to teach us what exactly happens when code words are used. Beliefs about words could fall on a wide spectrum of possible relations with the word, lexical meanings on one side of the spectrum and mere individual judgments on the other. As a result, this move cannot help Khoo to distinguish

---

[20]*ibid*, p. 51

his theory from Stanley's because it could be perfectly compatible with his view if we think of the judgment about the word as part of the conventional meaning of it.

The picture that Stanley draws for us about the formation of the conventional meaning is consistent with this reading. He says:

> [T]he links between word and meaning are a matter of degree, vague, and negotiable. The word 'Madagascar' originally referred to part of the mainland of Africa, but, because of changing usage, came to refer to an island off the coast of Africa. We see the same possibilities for change and resisting change with the kinds of repeated associations that propaganda involves.[21]

As a result, depending on the extent to which we think that the racial judgment is associated with the code word, Khoo's theory might be compatible with a semantic reading of the mechanism of code words. Khoo, however, has assumed that ruling out the possibility of conventional implicature using the non-cancellability test would buttress the non-semantic reading. Nevertheless, there are still other ways to understand the mechanism of code words which do not rely on inferences. In the next section I will elaborate on this point.

The second problem that arises from the uniqueness condition for Khoo's theory is that it cannot explain why racial judgment, among many other judgments that we have about code words, becomes salient when they are used in a *political discourse*. For instance, consider A who has a friend B who is white and lives in the inner city. A goes to a political rally and hears Mr. Ryan's political speech and hears him talking about the lack of the culture of work in inner cities. A has two different preexisting beliefs about the inner city:

1. B lives in the inner city.

---

2. The inner city is largely populated by African Americans.

So, according to my understanding of Khoo's theory, the speech could give rise to both of the following inferences:

1. My friend B lacks the culture of work.

2. African Americans lack the culture of work.

At least prima facie both inferences are equally lik ely to be drawn. But intuitively this speech does not evoke any resentment in A towards his friend B. Or even if his friend happens to be a hardworking person, he does not feel any contradiction hearing the speech. The morale of this discussion is that the theory is not able to explain the data.

Problems regarding the non-necessitation condition: Khoo's theory also cannot explain why people do not draw racial inferences in all contexts. That is, if people have racial beliefs about code words, why code words do not prime racial resentment in some contexts. For instance, the use of the code word in the graduation party does not result in racial inferences but Mr. Ryan's speech, according to Khoo's theory, gives rise to racial inferences. There is something special about political discourse that makes it fertile for racial effects. To shed more light on this issue, let's consider this case:

Imagine E is a fluent speaker of English who was living in Europe his whole life. Then, he moves to the US and meets A and B. He attends the same political rally that A did in the last case. However, hearing Ryan's speech about inner cities does not have the same effect on E as it does on A, even though he is in the same context and has the same familiarity with English as A does.

Khoo's theory supplemented with fragmentation would predict that E should draw the same racial inference because he, like A, exactly knows what 'inner city' means

and he knows about the racial makeup of inner cities. However, it is unclear how the theory could explain why E does not draw racial inferences. He might want to say there is some inference that E is missing, but given the description of the case, it is not clear what inference that is.

2.3   Saul: Covert Perlocutionary Speech Act

Jennifer Saul's theory, in contrast to Stanley and Khoo's theories, does not use any kind of derogatory content, either conventionally implicated or inferred, in order to explain the racial effect of code words.[22] Her view is that code words "must be understood as a species of perlocutionary speech acts."[23] She defines perlocutionary speech act as "the act of making utterance with certain effects."[24] The basic idea of hers is that code words are utterances with certain effects. By using a code word in a speech, the speaker intends to produce certain effects in audiences' minds and this effect has nothing to do with the meaning of the code word. The racial effect of code words, according to her theory, could be compared with the perlocutionary effects of making one feel happy, jealous, or angry with your words.

In order to capture the features of code words, Saul introduces a new category of perlocutionary speech act: covert perlocutionary act. According to her, "[a] covert perlocutionary act is one that does not succeed if the intended perlocutionary effect is recognized as intended."[25] She thinks that code words "are very much a matter of intended

---

[22]Saul uses the term 'dogwhistle' in order to refer to the same phenomenon. However, due to the following reasons I used 'code word' instead. First, because both Stanley and Khoo use 'code word' and I wanted to have a consistent vocabulary. Second, I think we should save the term 'dogwhistle' in order to refer to the phenomenon in which one sends a secret message to an in-group while the out-group do not get the message at all. But the case of code words is more complicated than this because it affects all the audiences, even though people who might explicitly reject any racist belief.

[23]*ibid*, p. 376

[24]*ibid*, p. 376

[25]*ibid*, p. 377

effects on their audience."[26] By making this move, she tries to make her theory able to explain two conditions. First, since she ties the effects of code words to the intention of speakers which is not public, the theory can account for the deniability condition in terms of speakers' not clarifying their private intentions. Remember Mr. Ryan's saying "I don't have a raci st bone in my body." In addition, the theory is partly capable of addressing the non-necessitation condition because according to the definition that she gives for the covert perlocutionary act, it functions as long as the intention of the speaker is not recognized. However, when one starts to discuss the racial effects of code words, the intended effect becomes explicit and the negative effect gets nullified.

And here is Lee's Saul-like explanation of what Ryan did. Mr. Ryan wants to prime racial resentment in the audience. In doing so, he uses the word 'inner city', knowing that it has a certain negative racial effect as long as his intention is concealed. He knows that when he uses the code word with the intention to influence the audience's biases, it produces a psychological effect the consequence of which is priming their racial resentment. However, the theory has the following problems.

Problems regarding the uniqueness condition: Saul's theory does not give us an account as to why and by what mechanism code words have the racial effect. If she insists that to explain the perlocutionary effect she "need not to understand them as about propositions. Nor need [she] claim that they are added to the common ground, or in any way consciously available to their audience", [27] what would be the resources by which she can explain the effect of code words? If all that matters is the intention of the speaker, why can't the speaker satisfy his intention using words that are not even remotely related to race, say 'gobbledegook' or 'vote'?

She claims that deception falls under the same category as code words do, i.e.

---

[26] *ibid*, p. 377
[27] *ibid*, p. 377

covert perlocutionary act. She says, "[o]ne who deceives can usually succeed if their intention to deceive is not recognized." [28] However, taking a closer look at the case of deception motivates us to be skeptical about the claim that we do not need to explain the effect of code words in terms of their content. Consider lying, for instance. Although it is controversial that intention to deceive is a necessary condition for lying, it is not disputable that when there is an intention to deceive, it would not be enough for lying to be realized. In addition, the truth value (or the believed truth value) of the asserted proposition must be false (or believed to be false). If one wants to lie about a specific state of affairs in the world, she cannot only intend to do so. She must pick words in order to assert the proposition such that they make the truth value (or the believed truth value) of the proposition false. The same goes for code words. The theory must provide us with an account as to the role that the code word itself plays in realizing the racial effect, irrespective of the intention of the speaker.

Problems regarding the non-necessitation condition: The way that Saul deals with the non-necessitation is unsatisfactory. Why does the perlocutionary act disappear when a code word is used in a different context? Here is her answer: "[n]ot every utterance using a particular [code word] will be intended to have the same effect." [29] In other words, the reason for this phenomenon for her is that the speaker, uttering a sentence containing a code word, does not intend to produce the racial effect. However, it is not clear to me how to specify the relation between the intention and the effect such that the mental state of the hearer plays no role in producing the effect. My intuition about the cases of non-necessitation, discussed in the first section, is exactly the opposite.

To illustrate, imagine that in the case of the city planners. Even if the speaker intends to prime racial resentment in the hearers, the racial effect will not follow because the intention of the speaker is unrecognized by the hearer. The reason is that the context

---

[28] *ibid*, p. 377
[29] *ibid*, p. 377

does not prime hearers to take the speaker to be intending to convey any racist message. I think this phenomenon should be explained in terms of an interplay of the speaker, the context, and the hearer. An explanation merely in terms of speaker's intentions would not be sufficient. In the next section I will elaborate on this point.

CHAPTER 3

A NEW THEORY OF CODE WORDS: POLYSEMY

After examining the three major theories of code words and showing their
problems, now I turn to my own positive proposal. Let's start our discussion with the
concept of polysemy. A polysemous word is one that "has several but related distinct
meanings whereby the same relation holds between the meanings for a series of nouns."[1]
Polysemy is usually understood in contrast with homonymy. While different senses of
homonymous words are unrelated, like *the edge of the river* sense of 'bank' and *financial
institution* sense of it, different senses of a polysemous word are connected to each other.
For example, different senses of the word 'glass' in the sentence "I put a glass on the table"
and "I drank a glass of wine" are related to each other. In the former sentence, 'glass'
refers to a container of a liquid and in the latter it refers to a unit of liquid. What is
important about polysemy is that it tends to occur systematically in language. That is, the
same relation holds between different meanings of several words in the language. For
instance, the same relation that holds between two senses of 'glass', holds between the
senses of 'bottle', 'cup', etc. [2]

There are various patterns in a language that alter the meaning of a word. These
pattern can be totally arbitrary and dependent on a certain context. However, in the case
of polysemous words, the patterns are well-established and conventionalized in language.

---

[1]Dolling (forth). p. 1
[2]Dolling (forth), p. 1

One common and conventionalized pattern in language, for instance, is to use the name of a place in order to allude to the community who live there. For instance, when one says: "The city has decided to increase the taxes," what she means by 'city' is *the residents of the city*. This pattern is known as PLACE/COMMUNITY.[3] Similarly, one can use the name of a place to denote a salient feature of the community who live there. For instance, if the common belief about the people who live in California is that they are liberal, one might use 'California' as an adjective that denotes *liberal* as in: "she has a California mindset."

My claim is that 'inner city' is a polysemous word that has different senses: RACIAL (black neighborhood), ECONOMIC (poor neighborhood), GEOGRAPHICAL (inner part of the city), and CRIMINAL (high crime neighborhood). Inner city as a geographical area of the city had probably been the primary sense of this word. However, the socio-historical facts of the US resulted in a set of homogeneous features that most of the inner cities have in common, such as economic status, racial makeup, and crime rate. As a result, people produced stereotypes about inner cities as neighborhoods which are populated by poor and mostly black people, and a lot of criminals live there.[4] Then, following the PLACE/COMMUNITY pattern, speakers started to use 'inner city' in order to refer to the features of the community who live there. Now, the primary sense of the 'inner city' is not primary anymore. That is, one can call a poor neighborhood an inner city even though it is not the actual inner part of the city.

Using of 'inner city' to denote those features is repeated enough and conventionalized that we can think of 'inner city' as a polysemous word that has several senses. For example, in the case of the graduation party, discussed in the second section, by 'inner city' the speaker denotes the ECONOMIC sense, a poor neighborhood, when she

---

[3]Dolling (forth), p.4.

[4]The claim here is not that these are real facts about inner cities. The point is that they are the stereotypical features that people tend to attribute to inner cities. To see some critical views on these stereotypes, cf: https://www.nytimes.com/2016/10/12/upshot/actually-many-inner-cities-are-doing-great.html and https://www.good.is/articles/trump-black-americans-inner-cities.

says: "she is raised in an inner city and now she got her PhD." In this context, the GEOGRAPHICAL sense of 'inner city' is totally irrelevant. In contrast, in the case of Mr. Ryan's speech, 'inner city', presumably, denotes a RACIAL sense, black neighborhood, when he talks about the lack of the culture of work in inner cities.

These shifts in the meaning has happened for various reasons. In the case of the ECONOMIC sense, it seems that we use 'inner city' to denote a *poor neighborhood* for euphemistic purposes. In the case of the graduation party, for instance, it sounds more appropriate to say "she is raised in an inner city" than to say "she is raised in a poor neighborhood." In the case of RACIAL sense, which is mostly used in political discourse, things are more complicated. In American political discourse, on the one hand, we have a history of political parties that want to take advantage of people's racial biases against black people and immigrants to get more votes. On the other hand, in this discourse what Mendelberg describes as 'norms of racial equality' are in place.[5] The consequence of these norms is that explicit derogatory racial references are prohibited and using them comes at a cost. As a result of this prohibition, code words shoulder the responsibility of conveying the racial message in lieu of their explicit racial counterparts and 'inner city' is one of them.

Now the important question is this: how does a polysemous code word, 'inner city' for instance, denote the RACIAL sense in a context among the different senses that it has? I think the key to answer this question lies in the fact that 'inner city' is underspecified. Vicente defines underspecification of a polysemous word as follows:

Underspecification: the meaning of a polysemous term is an underspecified, abstract, and summary representation that encompasses and gives access to its different senses.[6]

Thus, to say that 'inner city' is underspecified means that it has a core meaning

---

[5]Mendelberg 2001, p. 67
[6]Vicente 2018. p. 952.

23

that encompasses all different senses that it has. This core underspecified meaning in the context becomes specified and gives access to one of the senses. For instance, when Ms. Lee hears Mr. Ryan's speech, she interprets the RACIAL sense of 'inner city' among other senses. The reason is that she has an idea of his socio-linguistic character, and also she is aware of the norms that govern American political discourse which prohibit explicit racial references.[7]

To recapitulate, Ms. Lee would have given the following explanation of Mr. Ryan's speech, had she read this paper. Mr. Ryan uses the word 'inner city' which is a polysemous word, i.e. it has different but related senses. One of the senses of this word, that is commonly used by racist people, is a neighborhood of black people. Given the features of the context that prohibits explicit derogatory racial references and Mr. Ryan's socio-linguistic character given by the history of using this term to refer to black people by him and his like-minded colleagues convince us to think that the most appropriate interpretation of 'inner city' in this context is its RACIAL sense.

Here is how this theory is able to explain different features of code words.

Negative racial effect: Similar to Stanley's theory, this theory explains this effect in terms of the conveyed content. However, one advantage of this theory is that in order to explain the effect of code words it does not need to assume that the conveyed content is the same for all the audience. Using this theory, we can admit that various audience from different backgrounds interpret the coded speech in different ways and still be able to explain the effect. To elaborate, after hearing the coded speech, some already racist people might exchange a furtive smile with each other, meaning that we know what he is talking about, some people who realize the intention behind the coded speech might get angry, and

---

[7]There seems to be other clues in the speech that favor the racial interpretation such as 'culture', 'generations of men', and not working. These are the common vocabulary of racist speeches that are not as commonly used in the speeches about poor people in general. Thus, the most coherent interpretation that she can make is that Mr. Ryan is alluding to black people.

some people do not even interpret the word in the RACIAL sense. However, the theory can explain that even the latter group might be affected by the coded speech. This is because of the sustained activation of different senses of a polysemous word which is shown by psycho-linguists studies.[8] That is, even if one does not interpret the code word in the RACIAL sense and does not have the derogatory racial content in her conscious level, the racial content of the code word gets activated in the unconscious level and produces the negative racial effect. My guess is that the sustain activation effect is much more in the case of a political speech because of the fact that politician intentionally try to be more ambiguous in these cases and it makes it more difficult for the audiences to be certain about their disambiguation of the coded speech.

Deniability: The theory is also able to explain the deniability condition. Since, the process of interpretation takes place in the hearers' mind, the speaker can repudiate the interpretation of the hearer claiming that she did not mean the RACIAL sense. Also, it seems that the conventional use of the code word for the RACIAL sense is different from the conventional use of it for the ECONOMIC sense. While the latter is an accepted convention for the euphemistic purposes, the former is more like a tacit convention for a purpose that is always denied by the people who have that purpose. As a result, we observe that sometimes speakers deny the existence of such a conventionalized sense.

Non-necessitation: Now it is obvious that in the case of the graduation party, the audiences interpret 'inner city' in the ECONOMIC sense and in the case of the city planners, the GEOGRAPHICAL sense is the dominant one. However, having the sustained activation of different senses of polysemous words in mind, it seems that even in those cases we should observe at least *some* degree of racial resentment which does not seem right. My conjecture is that when the context definitely favors one interpretation over the RACIAL interpretation, the RACIAL sense takes a shorter time to decay than the cases of

---

[8]MacGregor et all 2015. p. 138

political discourse in which the water is muddy and interpretations are being made with less certainty. However, this claim must be examined by conducting empirical research.

Uniqueness: Since, according to this theory, conventionalized use and repeated association are the bases for the code words' RACIAL sense, we can understand why other words that stand in the same relation with the same racial group do not generate the same effect. Homogeneous vocabulary of politicians, when it comes to conveying derogatory racial contents, helps them to establish and conventionalize the RACIAL sense of code words and get the best result thereafter. They might use other words that conversationally implicate the racial content, but, having the features of polysemous words in mind, using polysemous code words is their best bet.

CONCLUSION

In this paper, first, I offered a framework for thinking about code words. This framework provides four primary features of code words that, altogether, distinct them from other kinds of harmful speech. The negative racial effect of code words, the fact that this effect is not necessitated by every use of them, the speaker's ability to deny the intention to prime the racial effect, and code words' significantly higher success in priming the racial effect than its coextensive words are the features that I discussed here. Having these four primary qualities in mind, I turned to three major theories that are offered to explain this phenomenon, i.e. Stanley's semantic theory, and Khoo and Saul's pragmatic theories. I showed how each of these theories fall short of giving a thorough explanation of code words. Drawing on those lessons, in the last section, I suggested a new theory of code words that, I think, is fully able to capture all four features of code words. This theory suggests that code words are polysemous. That is, they have different but related senses one of which is the RACIAL sense that gives rise to the negative racial effect.

BIBLIOGRAPHY

[1] Dolling, J. (forthcoming). Systematic Polysemy. In Matthewson, L., Meier C., Rullmann, H. and Zimmermann, T.E. (Eds.) *The Blackwell companion to semantics.*

[2] Grice, Paul H. (1989) .*Studies in the Way of Words.* Cambridge, MA: Harvard University Press.

[3] Horwitz, J. and M. Peffley. (2005) "Playing the Race Card in the Post-Willie Horton Era: The Impact of Racialized Code Words on Support for Punitive Prison Policy". *The Public Opinion Quarterly*69: 1, 99-112.

[4] Khoo, Justin. (2017). "Code Words in Political Discourse". *Philosophical Topics.* VOL. 45, NO. 2, Fall 2017. pp. 33-64..

[5] MacGregor, L. J., Bouwsema, J., Klepousniotou, E. (2015). Sustained meaning activation for polysemous but not homonymous words: Evidence from EEG. *Neuropsychologia*, 68, 126–138.

[6] Mendelberg, Tali. (2001). *The Race Card: Campaign Strategy, Implicit Messages, and the Norm of Equality.* Princeton, NJ: Princeton University Press.

[7] Saul, Jennifer. (2018) "Dogwhitles, Political Manipulation, and Philosophy of Language". In *New Work on Speech Act.* eds: Daniel Fogal, Daniel W. Harris, and Matt Moss. Oxford, Oxford University Press. pp. 360-83.

[8] Stalnaker, Robert. (1998) On the Representation of Context. *Journal of Logic, Language, and Information.* Vol. 7, No. 1, Special Issue on Context in Linguistics and Artificial Intelligence. pp. 3-19.

[9] Stalnaker, Rober. (2002). "Common Ground", *Linguistic and Philosophy*, 25 (5-6),

pp. 701-21.

[10] Stanley, Jason. (2015). *How Propaganda Works*. Princeton, NJ: Princeton University Press.

[11] Tesler, Michael, and Sears, David. (2010). *Obama's Race: The 2008 Election and the Dream of a Post-Racial America*. Chicago, IL: Chicago University Press.

[12] Vicente, Agustin. (2018). Polysemy and Word Meaning: an Account of Lexical Meaning for Different Kinds of Content Words. *Philos Stud*175:947–968.

[13] White, Ismail 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. 339-54.