MIT Open Access Articles

Code Words in Political Discourse

Code words in political discourse

Justin Khoo
MIT

Abstract. I argue that code words like "inner city" do not semantically encode hidden or implicit meanings, and offer an account of how they nonetheless manage to bring about the surprising effects discussed in Mendelberg 2001, White 2007, and Stanley 2015 (among others).

Political discourse in America and countries throughout the world is full of duplicity: insincere promises, outright lies, misleading statements, and so on. Recently, a more subversive phenomenon has been noted and discussed in the social sciences---the use of code words to subvert norms of democratic deliberation.[1] Here is a recent example, from Donald Trump's *Time to Get Tough: Making America #1 Again*:

> "*If we keep on this path, if we reelect Barack Obama, the America we leave our kids and grandkids won't look like the America we were blessed to grow up in. The American Dream will be in hock. The shining city on the hill will start to look like an inner-city wreck.*" (Trump 2011: 4)

At least three code words appear in this quote: "hock," "the shining city on the hill," and "inner-city." The first is a term for putting an item up for pawn, thus evoking impoverished desperation. The second alludes to the Sermon on the Mount, and has been used, going back to the Puritan John Winthrop, as a metaphor for American exceptionalism. The third carries racial connotations: it brings to mind poor, crime-ridden, African American neighborhoods. Using these code words in close succession, Trump contrasts two outcomes for America: a white, safe, prosperous, exceptional, Christian, America, and an America dominated by poor, lazy, and criminal African Americans. Yet, none of this is explicit: on the face of it, all Trump has said is that a second Obama term will result in the American Dream being put up for sale, and that what was once a shining city on the hill will begin to look like a densely populated urban area.[2]

---

[1] Code words are sometimes called "dog whistles" because (so the received view goes) they involve sending a message that can only be heard by audience members with suitably sensitive ears. Richard Morin is credited with introducing the term. He writes, "Subtle changes in question-wording sometimes produce remarkably different results … researchers call this the 'Dog Whistle Effect': Respondents hear something in the question that researchers do not." (Morin 1998).

[2] Here are two more examples of dog whistles. In his 2002 State of the Union speech, George W. Bush said, "there's power, wonder-working power, in the goodness and idealism and faith of the American people." Here, the phrase "wonder-working power" is a reference to a line from the Christian hymn "There's Power in the Blood." Edsall & Edsall 1992 discuss the racialization of certain Democratic themes among white blue-collar workers throughout the 1980s. Citing research done by a Democratic think (the Analysis Group) about white blue-collar workers from Macomb County, MI, they conclude that "the Democratic message by 1984 was viewed by one sector of the white electorate---a crucial sector in terms of presidential votes---through what might be called the prism of race; traditional liberal messages were passed through the racial filter; the word 'fairness' was read in racial terms, even when no explicitly racial content was intended" (Edsall & Edsall 1992: 183-4). This was particularly devastating to the Democratic party because many of their party slogans at the time involved appeals to fairness. Furthermore, Edsall and Edsall argue that "the continuing vitality of this coded language---a language of 'groups,' 'taxes,' 'big government,' 'quotas,' 'reverse discrimination,' 'welfare,' and 'special interests'---became critical to the maintenance of the conservative presidential majority in the latter half of the 1980s." (Edsall & Edsall 1992: 214).

The standard thought is that politicians use code words because they stand to gain from conveying certain messages implicitly, rather than stating them explicitly. This is surprising. Donald Trump wants us to oppose an Obama second term, and his strategy to secure our opposition is to insinuate that an Obama second term would result in white safety and prosperity giving way to black crime and poverty. Why does he not just explicitly say as much? Tali Mendelberg proposes that using a code word allows the speaker and her audience to violate certain social norms while plausibly denying doing so (Mendelberg 2001) For instance, by not saying anything explicit about African Americans, Donald Trump opens up room to plausibly deny that his statement is racist. Furthermore, someone may believe Trump's statement and on that basis oppose an Obama second term, and yet have room to plausibly deny that their doing so is racist (or based on racial fears). Relatedly, Jason Stanley proposes that the use of code words erodes norms of reasonableness by excluding certain perspectives from democratic deliberation, ultimately paving the way for their favored policies to be adopted on undemocratic grounds (Stanley 2015).

The question I am interested in today is *how* the use of code words achieves these surprising discourse effects. Both Mendelberg and Stanley propose variations on the idea that code words encode as part of their meaning a hidden, or implicit, message, which the user of a code word communicates, along with some other explicit message. This idea keeps with the received view of dog whistles as devices of communicating secret messages to the suitably-trained ear.

I think this kind of view is on the wrong track: what we identify as "code words" do not encode additional, implicit, meanings. Put oxymoronically, code words are not *code* for anything.[3] Instead, someone using a code word exploits (intentionally or otherwise) their audience's stereotypical beliefs about what they are talking about, without explicitly communicating these beliefs. Thus, using a "code word" allows (or leads) the audience to draw additional inferences from the speech without it being clear that they are doing so---and this is what distinguishes coded speech from speech where the relevant stereotypical beliefs are explicitly asserted. In a slogan: code words don't work by being vehicles of implicit communication; they work by triggering inferences which they are not used to communicate.

Here is the plan for the rest of the paper. After isolating some documented effects of code words (§1), I introduce Mendelberg's proposal that such speech creates space for deniable norm-violations (§2). I argue that, to work, Mendelberg's theory must pair with a semantic theory of the kinds of expressions which can be used to speak in code. In §3 and §4, I argue that Mendelberg and Stanley's proposals about the semantics of various code words fail. Then, in §5, I offer a simple theory of code words on which they do not linguistically encode implicit meanings. I argue that this semantics helps us understand how the use of code words creates space for deniable norm-violations. In §6, I turn to Jason Stanley's proposal that coded speech undermines reasonableness in democratic deliberation. I argue that my simple theory of code words provides a better account of this effect of coded speech than Stanley's theory. In §7, I extend my minimal theory of code words to social meaning more generally, and finally, in §8, I discuss some practical upshots of my theory.

1 Fixing ideas: some effects of code words

To situate our discussion, let us start with two examples of code words. My hope is that the following examples will help the reader to grasp the relevant effects distinctive of uses of code words.

---

[3] Therefore, on my view, the expression "code word" is a misnomer---"code words" are distinguished by their effects in discourse, not their meanings. Nonetheless, since the terminology is already somewhat entrenched, I will continue to use the expression (without scare quotes) throughout the paper. Furthermore, I agree with my opponents that certain words are well-suited to bring about the "code word"-type effects in political discourses in the (present day) United States, and I mean to use the expression to pick out that class of words.

The first example of a code word is "inner city." I want to focus on a puzzling effect of the use of "inner city" documented by White 2007. White aimed to measure the impact of "inner city" on how predictive an individual's racist stereotypical beliefs are of his or her (reported) support for food stamps.[4] In the study, participants' level of racial prejudice was measured by asking them "how well the words 'hardworking,' 'intelligent,' and 'lazy' described either Blacks or Whites as a group" (White 2007: 353).[5] Then, participants were asked to read a fictitious news story about a debate between two congressmen. The participants were divided into four groups (*African American, inner city, poor, nonracial*), and each read a slightly different version of the story---for instance, in the excerpt below, the differences are in the italicized portion (and which version corresponds to the name of the group):[6]

> … programs like food stamps and Medicaid represent important safety-nets for many *African American* / ***inner city*** / *poor* / *American* families, keeping them from falling deeper into poverty.

Importantly, the changes were entirely made to either direct or indirect quotes from the Democratic congressman character, who was arguing against the Republican welfare reform bill (for instance, the excerpted part above was from a direct quote from the Democratic congressman). White 2007 then measured how predictive belief in racist stereotypes about African Americans (in particular, the degree to which they thought white people were harder working, more intelligent, and less lazy than African Americans) is of opinions about food stamp programs among white participants. The main finding of the paper is that only in the "inner city" condition does greater racial prejudice towards African Americans become a significantly stronger predictor of opposition to food stamp programs. In that condition, such racial prejudice "becomes an important factor in shaping [white participants'] level of support for increased spending, with the most resentful Whites [those most negatively prejudiced against African Americans] being less than half as supportive as the least resentful, all else constant." (White 2007: 347). In other words, when food stamp programs are described as benefitting "inner city" families, white racial prejudice towards African Americans is significantly more predictive of increased opposition to government spending for food stamp programs than when the same food stamp programs were described as benefitting "African American families," "poor families," or just "American families."[7] Most surprisingly, racially prejudiced individuals were more likely to oppose food stamp programs described as benefitting "inner city families" than when such programs were described as benefitting "African American families."[8]

---

[4] The research paradigm here grows out of Mendelberg 2001. Mendelberg and others found that pairing otherwise neutral descriptions of candidates or issues with images confirming racist stereotypes tends to make belief in racist stereotypes a better predictor of that individual's attitudes towards social policies favoring African Americans and towards the politicians described than when paired with neutral images (Valentino et al. 2002), counter-stereotypical images (Mendelberg 2001, Valentino et al. 2002), and when the the description of the issue involves explicit endorsement of such racist stereotypes (Mendelberg 2001; see also Hutchings & Jardina 2009, Mendelberg 2008 for more results).

[5] This measure of racial resentment follows Kinder & Sanders 1996, which is the standard way of unobtrusively measuring the degree of racial bias in an individual.

[6] I include the full differences in the Appendix.

[7] Interestingly, White also found that both white and black participants were significantly more likely to mention crime as one of the three most important issues facing the country today when in the *inner city* condition as opposed to the control condition. However, white participants were also significantly more likely to mention crime in the *African American* condition as well. Thus, beliefs about crime cannot be the only mediating factor in generating the main result.

[8] A related study adds additional support to the claim that "inner city" conveys something about African Americans. Hurwitz & Peffley 2005 showed that the mere insertion of "inner city" in a question about spending money on prisons on antipoverty programs was sufficient to make participants' racial prejudice against African Americans more predictive of higher support for government spending on prisons. Their question was (p. 102-3):

The second example is from outside of the lab. Mendelberg 2001 discusses the example of George H. W. Bush's 1988 presidential campaign's use of Willie Horton in a series of coded campaign ads. William Horton was a black man convicted of murder and sentenced to life in prison. However, then-Massachusetts governor (and later to be Bush's Democratic opponent for president) Michael Dukakis supported the furlough program which allowed Horton a supervised weekend outside of prison. During his furlough, Horton escaped and proceeded to assault a white couple in their home, raping the woman. Bush seized upon Horton as evidence that Dukakis was soft on crime, repeatedly mentioning the name "Willie Horton" in ads and speeches (the nickname "Willie" was not Horton's, who went by "William" even to friends and family[9]). Mendelberg shows that the use of the name (and story) started without explicit mention of race (although often the stories were accompanied with pictures of Horton, clearly showing that he was black), and this preceded a sharp rise in Bush's poll numbers. However, in October of that year, Jesse Jackson accused the Bush campaign of using the Horton story with racial intent. This message, which was taken up and discussed by the media, preceded a sharp decline in Bush's poll numbers (though he of course ultimately ended up winning the election). Mendelberg concludes that the ads involved a coded racial appeal by combining the image of William Horton (and the name "Willie Horton") with the explicit message about Dukakis being soft on crime. The coded message conveyed by the words and image seemed to be that electing Dukakis would lead to more violent black crime. Perhaps an even stronger conclusion may be warranted, which is that over the course of the campaign, the name "Willie Horton" itself became a political code word used to stoke fear of increased violent black crime under a Dukakis presidency.

2 Mendelberg's theory

We turn next to understanding why code words can be used to bring about these surprising effects. I begin with Tali Mendelberg's theory (Mendelberg 2001), which will serve as the backdrop for what follows. As I mentioned earlier, the basic idea behind Mendelberg's theory is that coded speech frames an issue in such a way to allow for deniable violations of certain social norms, norms which might otherwise disincentivize someone from acting in this way. Let's unpack her theory by focusing on how it applies to the code word "inner city." To keep things simple, suppose Mr. Trump in fact said,

(1) If we re-elect Obama, America will start to look like an inner city wreck.

Mendelberg first observes that, throughout American society, there are prevailing norms of racial equality. That is, there are significant social, political, and economic costs to being perceived as racist. As such, people do not want to be thought of, and resent being called, racist---indeed, they do not even want to think of themselves as racist.[10] As such, if Mr. Trump had instead said,

---

Some people want to increase spending for new prisons to lock up violent [inner city/xxx] criminals. Other people would rather spend this money for antipoverty programs to prevent crime. What about you? If you had to choose, would you rather see this money spent on building new prisons, or on antipoverty programs? Do you feel strongly or not very strongly about this?

They conclude that "When messages are framed in such a way to reinforce the relationship between a particular policy and a particular group, it becomes far more likely that individuals will evaluate the policy on the basis of their evaluations of the group." (Hurwitz & Peffley 2005: 109).

[9] As revealed by Horton in an interview with *The Nation*: Elliot 1993.

[10] This is not the same as saying that there is no racism in American society---far from it. The point is that a general desire to be seen as adhering to norms of racial equality demands that such racism now be covert. Even Donald Trump recognizes this---on June 11, 2016, he tweeted in response to Mitt Romney's calling him a racist, "Mitt

(2) If we re-elect Obama, America will come to be dominated by poor, lazy, and criminal black people.

he would run the risk of alienating people from his cause. Such people may be sympathetic to Trump's message, but they also do not want their opposition to an Obama second term to be perceived (by others, and themselves) as racist. The problem with the overt message (2) is that it does not leave any room for someone to agree with it (outwardly, in their behavior, or inwardly, in their heart, so to speak) without violating norms of racial equality. By contrast, the coded racial appeal (1) leaves it unclear what Donald Trump wants his audience to infer---(2), or rather some principled concern about economics and crime? As a result, an audience member might agree with Trump's statement because she agrees with (2) or because she shares the principled concerns about economics and crime. By using racial code words, Trump makes space for the latter option as a possible interpretive option, and this allows Trump and his followers to plausibly deny the charge of racism ("Racist? Who, me? I'm just concerned about the spread of crime and poverty"). And it is this feature that explains why the use of the code word achieves a result (opposition to Obama on racial grounds), that could not be achieved by explicitly stating its coded message.[11]

We may state the theory generally as follows. There is a social norm **N** that governs the behavior of people in some community. As such, people in that community tend to conform their behavior to **N**---at the very least, there are social incentives to appear as if they conform to **N**. There is a code word 'C' that is ambiguous; it may mean F or it may mean G. Suppose some particular policy (or candidate) x is described using 'C' (someone says "x is C"). Furthermore, suppose that opposing x on the grounds that x is F would violate the norm **N**, but opposing x on the grounds that x is G would not. Then, the use of the code word frames x in such a way that allows someone who opposes x because x is F to plausibly deny having violating **N**. By contrast, explicitly stating that x is F does not allow for this possibility. As such, the code word opens up the possibility for people inclined to oppose x for F-reasons but who are also concerned about violating **N** to appear to have non-**N**-violating grounds for opposing x.[12]

Romney had his chance to beat a failed president but he choked like a dog. Now he calls me racist-but I am least racist person there is" (https://twitter.com/realDonaldTrump/status/741590381503086592).

[11] Coded speech also allows for plausible deniability of the speaker regarding the charge that they are talking about the alleged implicit message. To give one example, consider the following quote of Paul Ryan's (http://thinkprogress.org/economy/2014/03/12/3394871/ryan-poverty-inner-city/):

(i) 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 is a real culture problem here that has to be dealt with.

Paul Ryan here clearly aims for his message to resonate with speakers who believe that a central problem with welfare programs is that they reinforce a culture of dependency among African Americans (and black men in particular). Ryan was criticized on the grounds that his comments were racist against black people (http://www.crewof42.com/news/paul-ryan-inner-city-comments-had-nothing-to-do-whatsoever-with-race/). For instance, Barbara Lee said in response, "My colleague Congressman Ryan's comments about 'inner city' poverty are a thinly veiled racial attack and cannot be tolerated. 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.'" It is Ryan's response that is important for our purposes now: "This isn't a race based comment; it's a breakdown of families, it's rural poverty in rural areas, and talking about where poverty exists---there are no jobs and we have a breakdown of the family. This has nothing to do with race." Ryan here is attempting to cash in on the plausible deniability afforded him by the use of the code word---his strategy is to pin the problem on poverty rather than race.

[12] Michael Lynch argues that Donald Trump's unapologetic inconsistency benefits him by opening up an interpretive tension that hearers will typically resolve by interpreting him in whichever way confirms their own biases (Lynch 2016). However, contradicting oneself may also have the benefit of creating plausible deniability

On this theory, the code word 'C' must play a double role---publicly accepting the sentence "x is C" somehow puts one a position to oppose the policy x on *F*-grounds while denying that they are doing this (and insisting instead that their opposition to x is because x is *G*). There must be something about the meaning of 'C' that allows it to do this. For instance, if 'C' unambiguously means *F*, then anyone publicly accepting "x is C" (who understands the meaning of her words), and no other sentence about x, would not be in a position to oppose x and plausibly deny violating N---since opposing x on ^-grounds violates **N**. Thus, there must be some *semantic* feature of code words that allows them to be used in ways that affords plausible deniability of this kind.

In the next three sections, I will discuss three possible views about the meanings of code words like "inner city," which could in principle account for their ability to create space for deniable norm-violations. The first view is that code words are ambiguous (i.e., that 'C' is ambiguous between meaning *F* or *G*), the second is that code words have two dimensions to their meaning (i.e., that 'C' has *G* as its *at-issue* meaning and *F* as its *not-at-issue* meaning), and the third is that code words only have the non-norm-violating meaning (i.e., that 'C' means only *G*).

3 Ambiguity

According to the ambiguity theory, code words have at least two meanings; in the case of racial code words, one meaning is racial and the other is non-racial (I won't distinguish between polysemy and homophony accounts here). Consider the following use of "inner city":

(3) Food stamp programs help many inner city families.

According to this theory, "inner city" has two meanings---it may mean either "densely populated, high crime, urban area," or may mean "poor African American." Therefore, when someone says (3) they may mean either *Nonracial* or *Racial*:

- *Nonracial*: Food stamp programs help many families living in densely populated, high crime, urban areas.
- *Racial*: Food stamp programs help many poor African American families.

Someone hearing an utterance of (3), then, may accept it on its *Racial*-interpretation, and thus come to oppose food stamp programs on that basis, while also recognizing that there is an alternative interpretation of the utterance and hence their opposition. As such, they will believe that their opposition to the food stamp program has an outwardly ambiguous basis---others will not be sure whether they oppose the program because it helps African Americans or because it helps families living in densely populated, high crime, urban areas.[13] Thus, the ambiguity in "inner city" leads to an ambiguity in how to interpret the hearer's behavior, opening up space for a deniable violation of the norms of racial equality.[14]

---

among your supporters. If you advocate for a policy on racist grounds and then state that you support the policy but are not racist, denying your earlier claim, hearers are free to publicly interpret your *true* beliefs as the latter, even though they really support you because of your commitment to the former. Thus, by contradicting himself, Trump invites "bad faith" among his supporters---they are invited to support him because of his racist views, but also tell themselves and others that's not why they support him.

[13] The same rational reconstruction may also be put in entirely first-personal terms. The agent may not recognize or admit (to him or herself) his or her own basis for opposition to the food stamp program. In this way, the ambiguity drives self-deception

[14] The following quotes suggest to me that Tali Mendelberg favors this kind of theory of code words. She writes that an "implicitly racial appeal [is] one that contains a recognizable---if subtle---racial reference, most easily through

While this explanation seems adequate, I think there are good reasons to think that "inner city" (and other code words like "wonder-working power" or "illegal immigrant") are not ambiguous. Take the adjective "funny," for instance, which in English is genuinely ambiguous: it can mean either "humorous" or "strange." This property of "funny" is what allows for the possibility of predicating "funny" of someone while also saying they are "not humorous" or "not strange" without contradicting ourselves (cf. Zwicky & Sadock 1975, Sennet 2011):

(4) Smith is a funny man who is not humorous.

If "inner city" were ambiguous, then we should likewise be able to say things like the following without contradicting ourselves:

(5) #Smith is an inner city pastor who is from, works, and lives, in the suburbs.

But (5) strikes me as a contradiction, which suggests that "inner city pastor" can only mean "pastor from the inner city" or "pastor who works in the inner city"---it does not mean "pastor who is poor and African American."[15]

Let us look at another diagnostic to see the difference between "inner city" and genuinely ambiguous words. If *n* is genuinely ambiguous, it should be possible to fix the facts of the world, then find two contexts, one in which an utterance of *f(n)* is true and one in which an utterance of *~f(n)* is true (where *f(n)* is a declarative sentence containing *n* in a non-intensional position, and the two contexts do not affect the interpretation of any other expressions in *f(n)*).[16] So, take "bat" for instance. Suppose that Sue brought one baseball bat to the park and no chiropterans. Now consider the following two contexts:

(6) **Context 1**:
    A: I need a baseball bat. Did anyone bring one?
    B: Sue brought a bat. (True)

(7) **Context 2**:
    A: I am looking for a pet chiropteran. Did anyone bring one?
    B: Sue didn't bring a bat. (True)

Indeed, as evidenced by the difference in truth value in B's utterances, "bat" is genuinely ambiguous. Next, suppose that the densely-populated, high-crime urban core of City X is mostly populated by white

visual references" (Mendelberg 2001: 11), and that "Ambiguously racial messages activate racial predispositions but circumvent conflicting considerations such as a commitment to the norm of equality. Racial predispositions remain ready for use later on when voters form and express opinions about racial matters and when they choose among candidates" (Mendelberg 2001: 126). However, in personal communication, Mendelberg suggests that her view is actually a simpler, conjunctive theory, on which "inner city" means roughly "a densely populated, high crime, urban area, mostly populated by African Americans." This view is closer in spirit to Jason Stanley's multi-dimensionality theory, and is subject to the same criticisms as the latter (see §4).

[15] Strictly speaking, what matters here are the intuitions of individuals like those in White 2007's study, who were disposed to be racially prejudiced towards African Americans and act on that prejudice in the "inner city" condition but not "African American" condition. I am proceeding here on the reasonable assumption that we speak the same language as such participants (or at least a language similar enough in these respects), and hence can export conclusions about their meaning of "inner city" from our intuitions about it. Of course, more work is needed to justify this assumption.

[16] This test does distinguish ambiguous expressions from context-sensitive expressions; the latter will also past this test.

people, and that its suburbs are mainly populated by poor African Americans. Suppose that, yesterday, there was one arrest made for a crime committed in the city's densely-populated, high-crime urban core. Consider the following two contexts:

(8) **Context 1**:
    A: I need to know how many arrests were made in the densely-populated, high-crime urban core of the city yesterday.
    B: We made one inner city arrest. (True)

(9) **Context 2**:
     A: I need to know how many arrests were made in predominantly poor African American neighborhoods yesterday.
    B: We made no inner city arrests. (False)

Unlike with "bat" above, B's utterance is true only in the first context. This suggests that "inner city" is not genuinely ambiguous (with one of its interpretations being "African American neighborhood"). If it were, it should be natural to interpret B's utterance as true in **Context 2** (compare "bat"), but it seems that this is not the normal way of interpreting it.

Admittedly, it does seem possible to use words like "inner city" metaphorically to mean "urban African American," as in examples like:

(10) Smith has an inner city sense of style.

However, examples like this should not be taken as evidence that "inner city" is ambiguous. Almost any noun can be used as an adjective metaphorically given the right context. In an episode of the *West Wing*, one of the characters describes another as having a "New York sense of humor" meaning by this a Jewish sense of humor. This is an instance of synecdoche, and part of makes it successful is that there is no salient New York regional style of humor in the context (maybe Saturday Night Live, but what would that be?), so participants are left to infer the speaker's intended meaning from the context, and in particular, the obvious Jewish ethnicity of the addressee.[17] Furthermore, there is reason to think that it is not metaphorical uses of racial code words like "inner city" that explain the puzzling effects detailed by Mendelberg and White above. Such a theory predicts that the effects of code words should disappear in contexts in which the speaker is clearly intending to speak literally---but one such context is that of the congressman's speech in White 2007's vignette! This tells against the proposal that code words achieve their effects by being used metaphorically.

4 Multi-dimensionality

In his book on propaganda, Jason Stanley offers a more sophisticated view of code words on which they have two dimensions of conventional meaning: an *at-issue* content and a *not-at-issue* content (Stanley 2015). Stanley argues that this multi-dimensional theory of the meanings of code words helps to explain both their function in Mendelberg's theory, as well as explain how code words allow people to make prima facie *reasonable proposals* (those which are acceptable from the perspective of each person in the community) but which in fact serve to undermine "the capacity or the willingness to produce or be swayed by reasonable proposals" (Stanley 2015: 137-8; see also p. 117 and p. 134). We will come back to

---

[17] Consider the following variations, where someone says "You have a New York sense of style" to (a) a flashily-dressed man in a suit, or (b) an African American in Rocawear, or (c) an Orthodox Jew wearing a yarmulke and a tzitzit. Thanks to Alex Kane for pointing me to the *West Wing* example.

Stanley's claims about undermining reasonableness in §6. For now, I want to focus on how his theory may help us understand how code words make space for deniable norm-violations.

To begin evaluating Stanley's theory, we need to first understand the distinction between *at-issue* and *not-at-issue* content.[18] The distinction is between the components of speaker meaning (what someone means when they utter a sentence, in Grice 1957's sense), or utterance content, that are put forward as the *main point* of the speaker's utterance and those which are *backgrounded*:

- Not-at-issue content is *backgrounded*: A speaker (generally) expects the *not-at-issue* content of her utterance to already be accepted by her interlocutors (or otherwise be accepted by her interlocutors without discussion).
- By contrast, *at-issue* content is *foregrounded*: a speaker (generally) proposes that the *at-issue* content of her utterance be accepted by her interlocutors.
- A mark of this distinction is that *at-issue* content is (generally) available for direct denial, while *not-at-issue* content is (generally) not available for direct denial.

As an example, consider the sentence (11):

(11) John stopped smoking.

Someone uttering (11) commits him or herself to two propositions: that John no longer smokes, and that John used to smoke. The latter is a presupposition of the sentence triggered by the word "stop" and is its *not-at-issue* content, whereas the former is its *at-issue* content:

"John stopped smoking" >>
    a. **At-issue**: John does not smoke.
    b. **Not-at-issue**: John used to smoke.

The speaker uttering (11) presupposes that John used to smoke, and thus expects that her audience already believes this, or will come to believe it without discussion (this process is called "accommodation"; see Lewise 1979, Stalnaker 2002, von Fintel 2008). After all, that John used to smoke is not the main point of the speaker's utterance---what she is putting forward for consideration is the *at-issue* content that John does not smoke. Illustrating the second property of *not-at-issue* content, consider the following discourse:

(12)   A: John stopped smoking.
     B: That's false. / No he didn't (stop / #ever smoke).

Here, B's denial most naturally only targets that John stopped smoking---it is normally not understood as a challenge to the presupposition that John used to smoke. We would normally interpret B's saying "That's false" or "No" in response to A as claiming that John didn't stop smoking, not that he never smoked in the first place (that is why the interpretation of "No he didn't" where it means "No, he didn't ever smoke" is marked as odd). This is not to say that we cannot challenge *not-at-issue* content. The point is that we must use some more elaborate phrases to do so, as in (cf. von Fintel 2004):

(13)   A: John stopped smoking.
     B: Hold on---John never smoked in the first place! / Hey wait a minute---I didn't know
John smoked!

The proposal currently under consideration is that "inner city" encodes non-racial *at-issue* content and racial *not-at-issue* content. For example, a version of the view proposes that "inner city" has the following two components to its meaning:[19]

"Inner city" >>
        a. **At-issue**: A poor, densely populated, high crime, urban area.
        b. **Not-at-issue**: Those living in such areas are mostly African American.

Thus, according to this view, to say, "Food stamp programs help many inner city families" is to communicate both (the *at-issue* content) that food stamp programs help many families living in densely populated high crime urban areas and (the *not-at-issue* content) that such areas are mostly home to poor black people. Putting both components together, the communicated message is that food stamp programs help many families (who are mostly African American) living in poor, densely populated, high crime urban areas. Since the parenthetical part is *not-at-issue* and hence not available for direct denial, this explains why the following dialogue is inappropriate:

(14)   A: Food stamp programs help many inner city families.
                B: #No, poor, high crime, urban areas are mostly white.

The multi-dimensionality theory creates plausible deniability for those opposing the food stamp program because it benefits "inner city families." This is because stating that the food stamp program benefits "inner city families" conventionally communicates two propositions (that it benefits people in poor, densely populated, high crime urban areas and that most of those people are poor African Americans); in doing so, it opens up space for opposition to such programs on both grounds: (i) that its costs are not outweighed by its benefits to people living in poor, densely populated, high crime urban areas (regardless of their race), or (ii) that its costs are not outweighed by its benefits to poor African Americans. Furthermore, the *not-at-issue* content theory avoids the challenges facing the ambiguity proposal. The theory correctly predicts that (5) is contradictory, owing to its *at-issue* content being contradictory:

(5) #Smith is an inner city pastor who is from, works, and lives, in the suburbs.

Furthermore, the theory does not predict that there are two *at-issue* contents of "inner city" which may be brought out in different contexts, and hence it predicts the right results for **Context 2**:

(9) **Context 2**:
        A: I need to know how many arrests were made in predominantly African American neighborhoods yesterday.
        B: We made no inner city arrests. (False)

---

[19] Stanley 2015 proposes an alternative version of the *not-at-issue* theory in which it is part of the *not-at-issue* content of "inner city" that many people in the inner city are lazy and violent. Here is the relevant quote:

> For example, when a white US citizen is speaking to a black US citizen about the "problems in the inner city," there will be pressure, just to move the conversation forward, for the black citizen to say that she recognizes that many people in the inner city are in fact lazy and violent. So there will be pressure, just for conversation ease, to accept the stereotype of one's group, and of course then to personally distance oneself from that stereotype. Subordinate group members may be led to accept, however provisionally, the negative stereotype of their group, simply to enter smoothly into any conversation about their group with members of the dominating group. This is a consequence that follows straightforwardly from the linguistic mechanism involved. (Stanley 2015: 163)

Nonetheless, I think the multi-dimensionality theory faces problems of its own. A relevant property of *not-at-issue* content is that it is non-cancelable: someone cannot disavow commitment to the *not-at-issue* content of a sentence *S* that she utters merely by following up her utterance by asserting the negation of that content.[20] Take (11) and notice that one cannot disavow commitment to the claim that John used to smoke once one has uttered it:

(15) John stopped smoking. He never smoked at all.

This sentence is marked as odd because someone uttering it contradicts herself. However, it seems to me that code words do not have this property. Let's continue to focus on "inner city" (similar challenges will arise for Stanley's other example, "welfare"). Recall that we are supposing for now a version of the multi-dimensionality theory on which "inner city" has as its *not-at-issue* content that poor, densely populated, high crime, urban areas are primarily populated by African Americans. By the non-cancelability property of *not-at-issue* content, if this theory were correct, we expect utterances of sentences like (3) to non-cancellably commit the speaker to this content.

(3) Food stamp programs help many inner city families.

But this seems wrong. That is, it does not seem to be a contradiction to say:

(16) Food stamp programs help many inner city families, most of whom are white.

Unlike (15), this sentence does not strike us as self-contradictory. It may strike us as false, but it is clearly not inconsistent---we can imagine scenarios in which (16) would be true. It may be helpful to compare (16) with:

(17) #Food stamp programs help many inner city families, most of whom are wealthy suburbanites.

(17) here strikes us as a plain contradiction. Thus, the fact that (16) does not seem contradictory is evidence that "inner city" does not have as part of its conventional content that the poor urban areas are mostly populated by African Americans.

Now, Stanley might respond by noting that this test merely reveals that we have not yet found the right *not-at-issue* content for "inner city." However, other proposals, such as Stanley's own (that the *not-at-issue* content of "inner city" is that people in the inner city are lazy and violent; Stanley 2015: 163), also seem to fail the cancelability test:

(18) Food stamp programs help many inner city families, none of whom are lazy or violent.

Furthermore, recall that we are trying to explain the effects of code words like "inner city" on stoking racial prejudice against African Americans. To predict these effects on the multi-dimensionality theory, the *not-at-issue* content of "inner city" must be something about African Americans. However, it is hard

---

[20] This distinguishes conventionally encoded *not-at-issue* contents like presuppositions and conventional implicatures from conversational implicatures, which are cancellable (Grice 1989). A conversational implicature story would thus avoid the problem I raise here for Stanley's view, but it faces problems of its own. Conversational implicatures are supposed to be calculable on the basis of inferences about the speaker and various Maxims of conversation, yet it is unclear what such calculation would be in the case of "inner city." Furthermore, even if some uses of "inner city" carry conversational implicatures, as I discuss below in §5, such a theory seems otiose in light of the simple theory I develop.

to see what this *not-at-issue* content could be that would avoid the challenge raised by the consistency of (17).

The multi-dimensionality theory faces an important challenge---code words like "inner city" do not seem to carry racial *not-at-issue* contents. However, the failure of this theory points in the direction of a much simpler account of coded speech, which I turn to next.

5 A simple theory of code words

The problem facing both Mendelberg and Stanley's theories is that code words like "inner city" do not seem to encode anything about race as part of their meaning (either via ambiguity or multidimensionality); yet, given the results of White 2007's study, it seems that race is exactly the feature uses of "inner city" seems to make salient to participants deciding whether to support the food stamp program.[21] This may seem like a puzzle: how could "inner city" bring about this racial effect if it doesn't encode anything about race as part of its meaning?

In fact, as we will see, there is no puzzle here. Before we get there, though, let me articulate the view of the semantics of code words that my criticism of Mendelberg and Stanley has been building towards. On this view, code words carry no "implicit" meaning at all. This theory offers the most straightforward account of why code words afford plausible deniability as discussed above. (3) has only a nonracial meaning:

    (3) Food stamp programs help many inner city families.

Therefore, someone hearing this and thereby opposing such programs may plausibly deny that their doing so is racist---they may hold that their opposition is based on economic factors related to the urban poor.

Now, we come to the "puzzle" noted above: how does a view like this predict the racial effects of uses of "inner city"? Let me illustrate a minimal *inference-driven* account of the phenomena discussed in §2. Suppose Politician Z says:

    (A) The food stamp program will primarily benefit inner city Americans.
                                         Non-racial statement

According to the theory, Politician Z has not said (or even implicated, we may suppose---more on that below) anything about African Americans. He has just said that the program will primarily benefit Americans living in poor urban areas (regardless of race). However, hearers listening to Z's speech may believe:

    (B) The inner city is mostly populated by poor African Americans.
                                         Pre-existing belief

If they do believe this, then they will (consciously or non-consciously) infer from his speech that:

---

[21] Perhaps my arguments against "inner city" encoding something about race fail. Still, it may be that other racial code words do not semantically encode their racial connotations. Some examples where this seems particularly plausible are "illegal immigrant" and "underserved communities." The fact that there can be Canadian illegal immigrants and white underserved communities is evidence that these words do not encode their associated racial stereotypes (Latin Americans in the case of "illegal immigrant" and African Americans in the case of "underserved communities").

(C) The food stamp program will primarily benefit poor African Americans.
                                          Racial inference

Thus, though he does not communicate anything about race, since many hearers have pre-existing beliefs about the subject matter of what he *does* say, Politician Z's speech has the result that hearers will infer some racial belief about the policy from what he says. These hearers (and Politician Z himself) may thus oppose the policy on grounds that it will help poor African Americans, while plausibly denying that their opposition is racist: for instance, they may say that they oppose the program on the non-racial grounds that the urban poor already receive sufficient governmental assistance. If many of these hearers are motivated to avoid overt violations of norms of racial equality, then we expect to find greater opposition to the food stamp program among people in the "inner city" condition than from their counterparts in the "African American" condition. The reason is that "inner city" opens up a plausible non-racial source of opposition not easily available for those who hear the program described as benefitting African Americans.

Notice here that the politician may or may not have *conversationally implicated* that the food stamp program will primarily benefit poor African Americans---whether he does will depend on whether he intended his hearers to come to infer (C) on the basis of his saying (A). Crucially, though, the politician need not in any sense communicate that the food stamp program will primarily benefit poor African Americans for his speech to have these effects in his audience. Suppose, for instance, that the politician has no beliefs about the racial makeup of the inner city, and also does not believe that others think that the inner city is populated mostly by poor African Americans. The politician just says (A) because he believes that the food stamp program will primarily help people living in densely populated, high crime, urban areas. In this variation of the case, he unintentionally creates space for deniable norm-violations--- someone listening to his speech and who believes (B) will still come to believe that the food stamp program will primarily benefit poor African Americans, and thereby oppose it, and thus still have room to plausibly deny that their doing so is racist. Therefore, nothing in the account depends on the politician *meaning* (B) by saying (A). It seems then that nothing implicit need be communicated by a code word for a use of it to create space for plausible deniability regarding the violation of the norm of racial equality.[22]

I should emphasize that the beliefs which drive these inferences need not be about the subject matter of the word---they could be about the word itself. Thus, in the case of "wonder-working power," a hearer who recognizes the expression might come to believe that the speaker uttering it is a Christian.[23] Again, the important point is that this additional belief is inferred without being communicated. Such inferences are like those some of hearers may make about a speaker who, for example, speaks with a Southern accent---this person does not communicate in any sense that they are from the American South when they speak, although some of their hearers will infer this about them upon hearing them speak.

Consider another code word: "illegal immigrant." In the following example, we find it being used by the late Justice Antonin Scalia, who writes in a dissenting Supreme Court Opinion:

> "*Arizona bears the brunt of the country's illegal immigration problem. Its citizens feel themselves under siege by large numbers of illegal immigrants who invade their property, strain their social services, and even place their lives in jeopardy.*" (Arizona v. United States, 132, S. Ct. 2.492, 2522 (2012))

---

[22] We might characterize the view in Austin 1962's terminology: the crucial effects of code words are the result of perlocutionary effects of their uses, not their meanings or illocutionary force. Note that perlocutionary effects of a speech act may be unintentional (see Austin 1962: 106).

[23] This inference was found in Albertson 2015, and may be explained within the inference-driven theory by appeal to audience beliefs about the kind of people who use this expression.

Here, Scalia presents what may appear to some to be a principled reason to enact stronger border control measures (people's lands are *invaded*, their social services *strained*, and their lives *jeopardized* by undocumented immigrants). Yet, at the same time, for anyone who believes that illegal immigrants are typically Latin American, Scalia's message has as a consequence that it is Latin Americans who are invading property, straining services, and jeopardizing lives. Thus, as Ian Haney López notes, "Stressing illegality provides a way to seed racial fears without directly referencing race." (López 2014: 122-123). By not mentioning race, Scalia (or the many other people who frame border issues in terms of "illegal immigrants") opens up room for those supporting stronger border control measures on racist grounds to deny doing just that. Such individuals may instead insist their support is non-racial---they just want the letter of the law to be upheld.[24] Just as before, Scalia (or anyone using the code word "illegal immigrant") need not have intended to communicate anything about race: the inference goes through regardless because of the pre-existing stereotypical belief about undocumented immigrants.[25]

The take home message is that code words need not be code for anything to have "coded" effects. Here is what is happening, schematically. Take some code word 'C.' We distinguish the pre-existing 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.

The speaker asserts the **Explicit Statement** to an audience who already believes the **Existing Belief**. Believing the speaker, then, the audience comes to believe **Inferred** (or, more cautiously, the audience provisionally infers **Inferred** from **Explicit Statement** and **Existing Belief** in order to evaluate **Explicit Statement**). As I noted above, this inference need not be consciously drawn---a hearer may come to unconsciously believe **Inferred** on the basis of accepting **Explicit Statement** and **Existing Belief** without this arising to the level of introspective awareness. But now notice that there is room for plausible deniability. The audience may have some attitude towards x because it is C or because it is R, and this allows them to plausibly deny that their attitude is based on R.

I want to pause to discuss an important complication facing this simple theory of code words. It may seem that the theory predicts that the effects of a code word 'C' should persist across substitutions of an expression 'D' that has the same extension as 'C.' The reason this may seem to be so is that, according to the simple theory, it is beliefs about Cs that are doing the work, not anything about the meaning of the word 'C.' Suppose we change the congressman's speech from the *inner city* condition of White 2007's study to read:

---

[24] Interestingly, as López 2014 points out, "illegal" is actually a misnomer: "crossing into or remaining in the United States without proper authorization is not a crime, but rather a civil matter. Thus, contrary to Scalia's preference, the Court majority eschewed the term 'illegal alien.'" (122).

[25] Here are two more examples, for good measure. The first is "states' rights," which means "the rights of of self-governance afforded to individual states." This term was first popularized by Barry Goldwater in his opposition to the 1964 Civil Rights Act. By framing the issue in this way, Goldwater could clothe his racist commitment to segregation in the guise of a principled commitment to preserving state autonomy and resisting the illegitimate overreaching of the federal government. The second example is the expression "voter ID laws," which clearly only means "the laws which require voters to have government issued IDs." Someone who believes that voter ID laws would disproportionately hurt Democratic voter turnout (especially among African American and Latino voters) may then be in favor of voter ID laws for this very reason (which would violate norms of racial equality), and yet plausibly deny this by instead citing concern about voter fraud.

> … programs like food stamps and Medicaid represent important safety-nets for many families living in densely populated, high crime, urban areas, keeping them from falling deeper into poverty.

Assume for now that "inner city" and "densely populated, high crime, urban areas" have the same extension. It is possible that we would find the same pattern connecting racial prejudice towards black people and opposition to food stamp programs in this condition as we do in the *inner city* condition. If we find the same pattern, this would suggest that it is not essential to use the phrase "inner city" to bring racial beliefs about the inner city to mind. However, if instead we find that this condition patterns more like the *poor* condition, this would suggest that "inner city" is doing something that brings to mind racial beliefs about the inner city that "densely populated, high crime, urban areas" does not do.[26] We might think that such a result would disconfirm the simple theory discussed above.

However, this is not right. All that such a result would suggest is that the simple theory needs supplementation in some way. The reason that this by itself would not falsify the simple theory has to do with the possibility of fragmentation. A rational individual may believe x is F when thinking of x one way, and believe that x is not F when thinking of x in a different way. Here is an example from Elga & Rayo manuscript (following Lewis 1982):

> "*Whenever Andy travels to Leicester Square on foot, he treats Charing Cross Road as though it runs east-west. But whenever he travels to Leicester Square by train, he treats Charing Cross Road as though it runs north-south. And this is so even though he is confident that there is only one [Charing Cross Road] near Leicester Square.*"

When Andy considers Charing Cross Road in the "traveling by foot" way, he thinks it runs east-west (at least, he is disposed to act in a way revealing such a belief); yet when Andy considers Charing Cross Road in the "traveling to Leicester by train" way, he thinks it runs north-south. Elga & Rayo manuscript argue that there may be no elicitation-condition-invariant way of specifying what Andy believes about Charing Cross Road (cf. Lewis 1982, Stalnaker 1991, 1999, Schwitzgebel 2010). Similarly, it may be that certain people tend to believe that densely populated urban areas are mostly populated by African Americans when they think of them as inner cities, but don't tend to believe this when they think of these areas as densely populated urban areas. If this is right, then the simple theory will not predict the same pattern connecting racial prejudice towards African Americans and opposition to the food stamp programs in the above variation as compared with the *inner city* condition.

Finally, I want to contrast the above *inference-driven* theory with an *association-driven* theory of code words (the two theories share in common a simple semantics for code words on which they are not code for anything---i.e., both theories agree that "inner city" has no racial meaning in any sense). However, on the association-driven theory, it is not belief about the racial makeup of the inner city, together with racial prejudice towards African Americans, that explains its racialized effects. Rather, it is an association between "inner city" (or the concept INNER CITY) and the concept of AFRICAN AMERICAN (or maybe just RACE) which then primes racist beliefs and prejudices (as in Greenwald & Banaji 1995, Greenwald et al. 1998, Banaji 2002). These are very different mechanisms, and, as such, they predict different effects of code words (cf. Mandelbaum 2015, who helpfully distinguishes these two kinds of theories of implicit bias). For instance, the simple theory predicts the effects of code words like "inner

---

[26] There is a potential confound here. It would still be possible that the reason "densely populated high crime urban area" does not activate racial predispositions to the extent "inner city" does is that some people believe the two expressions have different extensions. A possible way to control for this possibility may be to ask participants in the study whether an area could be "inner city" even if it isn't "densely populated with a high crime rate" and vice versa.

city" should be restricted to judgments about the inner city and propositions inferentially related to beliefs about the inner city, while the association theory predicts that we should find such effects extending beyond judgments about the inner city (and subject matters inferentially related).

A further distinguishing mark of the association theory is that it, but not the simple theory, should predict that the relevant stereotypes can be primed by mentioned tokens of "inner city." Thus, for instance, if it were associations between "inner city" and the concept RACE that primed rational resentment in experiments such as White's, we would expect a similar pattern of data in a condition in which the Democratic congressman doesn't use "inner city" but merely mentions it:

> … my colleagues like to talk about the "inner city" to scare voters, but I think that is just dog-whistling. Programs like food stamps and Medicaid represent important safety-nets for many families, keeping them from falling deeper into poverty.

Furthermore, there is good reason to think that if associations of "inner city" are priming racial prejudice and generating increased opposition to food stamp programs, such associations should be able to be primed subliminally. Thus, we might flash "inner city" on the screen when the neutral version of the transcript is being read. In this condition, the association theory ought to predict the same result of increased opposition to food stamp programs among participants who are racially prejudiced against African Americans. Evidence that this is a viable way to test implicit associations comes from an actual campaign ad run during the 2000 election. The ad, sometimes called the "RATS" commercial, was run by the RNC. During the commercial's discussion of Al Gore's Medicare drug plan, the word "RATS" was subliminally flashed on the screen for one-thirtieth of a second. In a controlled experiment, Weinberger & Wester 2008 found that merely flashing "RATS" on the screen during the ad led to a stronger negative evaluation of the candidate being attacked (when compared with "STAR," "ARAB," and no subliminal stimulus).[27] For more discussion of subliminal priming see Bar & Biederman 1998, Abrams et al. 2002, Stewart 2008, Gibson & Zielaskowski 2013.

At this point, it should be clear that we do not yet have enough empirical evidence to decide whether the mechanisms activating racial predispositions are cognitive or associational. Thus, although I am adopting and defending the simple theory in this paper, I take this commitment to be provisional and subject to additional evidence. I think this is an exciting area for future research in the area of code words, one which will require the collaboration of experts in linguistics, philosophy, political science, and psychology. In the rest of this paper, I want to try to make a little more progress from the armchair, by considering Jason Stanley's proposal that code words undermine norms of reasonableness. I will argue that the undermining power of code words actually tells against Stanley's theory that they are multi-dimensional, and hence counts in favor of the simple theory.

6 Undermining reasonableness

In his book on propaganda (Stanley 2015), Jason Stanley aims to explain how seemingly reasonable proposals can in fact erode norms of reasonableness of public discourse. Stanley proposes that "To preserve the character of democratic deliberation, those deliberating in formal and informal debate over policy are subject to a norm of reasonableness, which requires them to take the perspectives of others into account" (Stanley 2015: 124).[28] To follow the norm of reasonableness governing public speech is "to take one's proposals to be accountable to everyone in the community. A reasonable person only acts in ways

---

[27] That "ARAB" had no significant effect may seem surprising, but this may be partly explained by the fact that the study was run in 2000, prior to the events of September 11, 2001.

[28] Stanley cites Rawls 2005 as a precedent for this idea.

that would be acceptable from every perspective; the reasonable person takes herself to be accountable to all her fellow citizens … A community governed by the normative ideal of reasonableness is one in which citizens have mutual respect for everyone else in the community and take their actions to be accountable to everyone else in the community" (Stanley 2015: 108). So, for example,

> "*You may decide to become Christian; I may decide to become a Scientologist; and a third friend may finally settle on atheism. These are all reasonable albeit incompatible paths. By imposing an ideal of reasonableness, Rawls is requiring reasons to not be drawn from the differing doctrines fellow citizens hold as a consequence of the decisions they made that formed a legitimate life path. On this view, it is not permissible for you to draw on your Christian beliefs in public debate, because, if you are reasonable, you are aware that Christian doctrine is not reasonable from my perspective.*" (Stanley 2015: 107)

Thus, the norm of reasonableness demands that the reasons publicly cited for or against a particular policy must be ones that are accepted as such from the perspective of each member of the community. In other words, we would violate the norm of reasonableness if we cite as a reason for or against some policy one that is not so-accepted by some perspective in the community.

Presumably, the norm of reasonableness will be restricted to only those perspectives in the community which are worthy of respect---that is, which have a legitimate claim to be taken into account in democratic deliberation.[29] This allows for some flexibility in how the norm will be interpreted within a community. Compare three communities, C_1, C_2, C_3, each of which contains members adhering to perspectives A and B and each of which is governed by the norm of reasonableness. Each member of C_1 agrees that perspectives A and B are worthy of respect. Then, there will be social, economic, and political costs in C_1 to publicly supporting (or opposing) policies on grounds that are not accepted as such by either perspective A or B. Compare this community with C_2, of which some members think that A and B are both worthy of respect, while other members think that only A is worthy of respect. In C_2, there will be social, economic, and political costs to publicly supporting (or opposing) policies on grounds not accepted as such by perspective B, but they will be lessened when compared with similar behavior in C_1 (since in C_2 not everyone agrees B is worthy of respect). Finally, compare both with community C_3, in which everyone only thinks perspective A is worthy of respect. In C_3, there won't be social, economic, and political costs to publicly endorsing (or opposing) policies on grounds that are not accepted as such by B.

Notice that this talk of norms is deployed descriptively, as it was for Mendelberg and the norm of racial equality. For all that has been said, it may be that both perspectives A and B are worthy of respect. Suppose that's right. Then the norm of reasonableness is incorrectly deployed in C_2 and C_3 (I take it this is what Stanley means when he says the norm is "eroded"). When this happens, democratic deliberation is undermined because speakers may publicly cite parochial reasons for some policy (ones that are not accepted as such by certain legitimate perspectives in the community), thus allowing democratically illegitimate reasons for adopting that policy into the public discussion. Ultimately, the community may end up adopting policies which are democratically illegitimate (as in policies voted for on the basis of deception).

For Stanley, propaganda is the mechanism by which one may seem to make a reasonable contribution to a discourse, while in fact framing "the debate in such a way as to exclude the perspective of a targeted group" by creating "flawed ideological beliefs to the effect that the perspectives of a designated group are not worthy of reasonable consideration" (Stanley 2015: 129). In other words, propaganda is a mechanism

---

[29] Evidence that this is how Stanley is thinking about things comes from his talk of "expanding the domain of respect and empathy" and "excluding individuals from the domain of respect" (Stanley 2015: 114, 116, 149).

by which apparently reasonable proposals erode the norm of reasonableness within a community---moving the community from one like C_1 towards one like C_2 or C_3. Stanley appeals to the not-at-issue content theory to explain how such propaganda works. Code words which have a non-racial at-issue content but racial not-at-issue content allow speakers to make an apparently reasonable claim (owing to its at-issue content being reasonable) while at the same time framing the debate in such a way to exclude certain perspectives from being worthy of respect (owing to its not-at-issue content).

One of Stanley's examples is "welfare," which he claims has as its not-at-issue content that blacks are lazy (Stanley 2015: 138, 163; I think more plausible is the proposal that its not-at-issue content is that *poor* blacks are lazy---I'll assume this is Stanley's view in what follows). Supposing Stanley is right about this. We can now explain how uses of "welfare" may seem reasonable but may in fact erode the norm of reasonableness by excluding the perspective of poor African Americans as worthy of respect. Suppose a politician says,

(19) We need to reform our welfare system to put incentives in place for people to work hard.

According to Stanley, the speaker uttering (19) asserts its non-racial at-issue content, and also communicates the not-at-issue content that poor blacks are lazy. Since its non-racial at-issue content is reasonable, the contribution is felt to be reasonable. However, its not-at-issue content is expected to be already common ground, or otherwise accepted without discussion, and as such typically will be common ground after the speaker finishes talking (unless it is explicitly blocked). But the consequence of accepting this not-at-issue content will tend to decrease empathy for poor black people ("they are lazy, so they don't deserve government assistance") and thus bring about the belief that the perspective of a poor black person is not worthy of respect. As such, uses of "welfare" will tend to covertly erode reasonableness by lessening the community's sanctions against publicly supporting policies on grounds not acceptable from the perspective of poor black people.

We have already seen that "inner city" fails the cancelability test of not-at-issue meaning, and the same diagnostic shows that "welfare" also does not conventionally encode as its not-at-issue meaning any negative stereotype about African Americans and nor does it express an attitude of diminished empathy for poor black people:[30]

(20) Our welfare system provides needed services to many unfairly disadvantaged citizens. Every negative stereotype about poor black people is false.

(21) Our welfare system provides needed services to many unfairly disadvantaged citizens. I am deeply empathetic towards the difficult situation faced by poor black people.

Neither (20) nor (21) sounds contradictory. Yet, Stanley's multi-dimensionality theory predicts they should sound like contradictions, just as (15):

(15) #John stopped smoking. He never smoked at all.

A second problem with Stanley's theory is that it predicts that uttering a code word will all by itself erode the norm of reasonableness. As long as the code word's not-at-issue content projects out of the linguistic environment it is embedded within, non-vigilant audience members will come to accept the code word's not-at-issue content and as a result come to believe that some perspective is not worthy of respect. I have given some reason to doubt this view: code words like "welfare" and "inner city" don't seem to pass the

---

[30] Stanley seems neutral between these two possibilities, calling the former the "content model" and the latter the "expressive model" (Stanley 2015: 140).

usual tests of encoding conventional not-at-issue content. I want to bolster this claim now by contrasting two instances of "welfare," one which I think *does* have the effect of excluding certain perspectives while the other does not:

(22) We need to continue to fund our welfare system because it provides needed services to many unfairly disadvantaged citizens.

(23) We need to reform our welfare system to put incentives in place for people to work hard.

I think it is implausible that (22) has the effect of excluding any perspectives (including those of poor black people). Rather, (22) demands that we consider and respect the perspective of welfare recipients---it presupposes that many welfare recipients are unfairly disadvantaged and proposes that helping them is a reason to continue to fund the system. If this is right, then the use of "welfare" does not, all by itself, serve to exclude certain perspectives as being unworthy of respect.

By contrast, (23) proposes as a reason for welfare reform the need to incentivize welfare recipients to work. It presupposes that a significant number of welfare recipients could work hard but do not (they are lazy), and also that the current welfare system provides no incentive for them to work hard. As such, it frames the issue of welfare reform as one of preventing the unfair exploitation of a common good by lazy individuals. This seems to be a reason for reform that should be accepted as such by any perspective in the community that is worthy of respect. That is, given (23)'s presupposition, if a perspective did not accept it as such, we would think that perspective unduly parochial and hence not worthy of respect.

But now notice that an individual X who accepts the presupposition of (23) may, upon encountering someone Y who rejects (23), dismiss Y's perspective as not worthy of respect by reasoning that they must reject a reason for welfare reform that is accepted by every perspective worthy of respect. Of course, Y might reject (23) *not* because she thinks it is fair that lazy individuals exploit a common good, but rather because she disagrees with the framing of the issue as one of preventing the unfair exploitation of a common good by lazy individuals. However, unless X and Y are sufficiently able to engage critically with each other, X may not realize that this is where their disagreement lies. As such X will continue to simply dismiss Y's perspective as unworthy of respect. Thus, democratic deliberation is threatened---policies adopted on such grounds will not be reasonable from every respect-worthy perspective in the community, though many members of the community will not recognize this fact.

Examples like (22) and (23) suggest that it is not the use of code words that erodes the norm of reasonableness, but something else---in this case, the presuppositions triggered by the "because" and "in order to" clauses (in (22) this is just a "to"-clause). More examples should be discussed to further verify that this challenge to Stanley's proposal is on the right track, but the initial evidence is telling.

7 Social meanings

On my view, there is no semantically distinctive class of expressions which we might think of as "code words." Instead, what we have are various words which may be used in certain contexts to bring about certain surprising cognitive effects, owing to beliefs (or associations) people have about the subject matters of these words (or the words themselves). Thus, we would expect many words which are not normally considered "code words" in the ordinary sense to exhibit similar effects. This is a point where my view connects up with the phenomenon of social meaning.

The concept of social meaning comes to us from legal scholarship (see Lessig 1995, Kahan 1997, as well as Haslanger 2013). By the social meaning of an action or status, we mean its social significance to the members of a particular group. For instance, consider the different attitudes we have towards meat rather than flesh, and how the color pink seems to naturally pair with being a girl and blue with being a boy. The *social* aspect here is important: not only do we tend to infer that an infant dressed in pink is a girl, we tend to think that the parent has dressed their child in pink deliberately with the intention of conveying that information, and that the parent thinks that we think that, and so on.

Stanley also sees a connection with his theory of code words and social meaning, and suggests that his theory of not-at-issue content may help to elucidate the latter concept (see Stanley 2015: 157-8). Above, I gave some reasons to doubt that Stanley has articulated the correct theory of code words, and I want to note here that there is no clear way to extend his proposal to account for the social meaning of non-linguistic entities (like actions or images)---a point that Stanley concedes (Stanley 2015: 127). But even if we just focus on its application to the social meaning of linguistic expressions, Stanley's theory of social meaning seems to yield the wrong results. Take, for instance, the word "professor" which Stanley suggests has as its social meaning "someone with liberal political views, who is practically incompetent" (Stanley 2015: 169). However, this cannot be part of the linguistic meaning of "professor" in any sense (either its at-issue or not-at-issue content)---if it were, describing someone as a right-wing professor would be incoherent.

By contrast, the simple theory of code words that I defend above *does* extend naturally to the social meanings of linguistic expressions, as well as actions and statuses. The view is simple: the social meaning of *x* within a community is a cluster of shared (or overlapping) stereotypical beliefs members of that community take to be shared among them about *x*.[31] Although social meanings are not linguistic meanings, they are related in a particular way. When we have stereotypical beliefs about something, we by default imagine stereotype-congruent instances of that thing, and so if we believe these beliefs are shared, describing something as "F" will generate a default implicature that it is a stereotypical F (Levinson 2000). So, for instance, if I say, "I went for a run yesterday," I will by default implicate that I ran in shoes, even though this implicature is obviously cancellable: "I went for a run barefoot yesterday." In the case of "professor," we have many shared stereotypical beliefs about professors which may include that professors tend to be liberal and practically incompetent. By describing someone as a professor, then, I may implicate that they are liberal and practically incompetent. However, since the truth of the generic statement that professors are liberal is compatible with there being exceptions to this generality, it remains possible to coherently believe that professors are liberal and also believe that Smith is a right-wing professor. As with code words above, the social meaning is tied to "professor" by way of being a widespread stereotypical belief about professors (the subject matter of the word).

Let us take a moment to explore a particular example of social meaning discussed in Lessig 1995, the social meaning (or significance) of being a "military man." This example helps to illustrate another reason why social meaning is not linguistic meaning---in this case, by showing that when the social meaning of a certain status becomes ambiguous, intentions are insufficient to disambiguate, and the status loses

---

[31] This is view is simplified in two dimensions. For one, it is entirely cognitive. But, as Haslanger 2013 argues, social meaning extends beyond representation to include schemas which we use to "focus attention, coordinate expectations, sift evidence, and rationalize behavioral and emotional dispositions" (27). I am sympathetic to Haslanger's arguments, so a full account of social meaning should go beyond shared *beliefs*. Secondly, the simple view incorrectly classifies some folk beliefs of the subject matter (such as folk physics) as genuinely social meaning, and yet this seems an overgeneralization. A possible response is to restrict the beliefs to those which comprise our beliefs about the social world (i.e., our expectations about other people's beliefs about x), in which case the belief that speakers tend to follow Grice's Cooperative Principle would be part of the social meaning of being a speaker. I leave sorting out this matter for another day.

whatever social power it had. This is unlike linguistic meaning where we standardly defer to speaker intentions when interpreting what was meant by an utterance of an ambiguous word.

Lessig notes that, among many Americans (and perpetuated by US military recruitment efforts), the stereotype of a military man involves exemplifying various virtues, among them strength, discipline, dependability, and patriotism. Thus, to be perceived as a military man carries a certain social prestige---people will tend to believe that you are strong, disciplined, dependable, and patriotic. However, among many individuals for which being a military man has such a social meaning, homosexuals are seen as stereotypically weak, irresolute, and effeminate. From these premises, Lessig infers that if the military were to openly open its ranks "to homosexuals, [this] would ambiguate the social meaning of membership in the military, as well as ambiguate the meaning of being gay for those who hold the stereotypical view. If openly open to all males, the military could not preserve the image of being constituted exclusively by [men exemplifying virtues of strength, discipline, and heterosexuality]" (Lessig 1995: 988). Thus, to allow openly gay men in the military might result in military men no longer carrying the same social prestige. As such, for young men who "join the military in part to gain their perceived value of this association, there is a strong interest in preserving the image that the military presents. For them, part of the value in belonging to this military depends upon the preservation of this image." (Lessig 1995: 987). Its social meaning sufficiently ambiguated, being a military man would no longer have among its perceived benefits that people in your community believe you are strong, disciplined, dependable, and patriotic. Lessig suggests that a desire to preserve the social prestige of being a military man may be a crucial factor in explaining resistance to allowing openly gay men to serve in the military, and to understanding the appeal (among military leaders) of the "don't ask, don't tell" policy adopted in 1993.

Importantly for our purposes is Lessig's observation that resistance to allowing openly gay men to serve in the military may be grounded in the fact that doing so would ambiguate the social meaning of being a military man, which in turn would undermine the social prestige of the position. If, by contrast, social meaning were like linguistic meaning, then ambiguity would be a complication, but one which could be clarified by speaker intentions. However, it seems clear that social meaning does not work this way. Once the military allows openly gay men to serve, and this is commonly known, people may no longer believe that you are strong, disciplined, dependable, and patriotic merely because you serve in the military.[32] No amount of intending for people to believe otherwise will help. Obviously, a military man could intend for others to believe he is strong, disciplined, dependable, and patriotic on other grounds, such as by lifting large objects and keeping his promises. The point is just that *mere* allegiance with the military is no longer sufficient, and intending for it to be sufficient is a lost cause.

Social meaning just does not seem to be linguistic meaning. Perhaps it is closer to what Grice labeled "natural meaning." However, this term stands to mislead, on the grounds that many of our stereotypical beliefs do not track natural facts (nor do they purport to). Maybe a term less apt to confuse along either dimension is "social significance." This brings to mind the sense in which things (words, clothes, etc) may signify things within a community (race, status, political stance, etc) but remain neutral about how that signification relation works.

8 The way forward

The presence of code words in political discourse raises several important practical questions for concerned citizens. What might we do to inhibit these effects of code words? Can we prevent racism from

---

[32] I say "may" here because, once gay men are openly serving in the military, the stereotypical beliefs about gay men within the relevant community will probably change. In particular, it seems plausible that gay military men will also be seen to be strong and disciplined.

"going underground" in the face of norms of racial equality? And how should we combat "racism 2.0"---the kind of racism displayed by Donald Trump and his supporters in a reaction to the previous generation of coded racial appeals?

These are challenging and pressing questions, and I cannot hope to begin to answer them here. However, I think that my theory of code words may help to refine our strategy for resisting coded racial appeals. Mendelberg 2001 argues at length that coded racial appeals succeed to the degree to which it is not commonly accepted that they are racial appeals, and she recommends a strategy of explicitly "calling out" the appeal as racial.[33] However, Mendelberg also cautions that,

> "*Candidates who are challenged have increasingly been responding with an accusation in kind. Some candidates, when pressed on their use of racial appeals, argue in turn that it is not they but their challenger who is injecting race into the campaign. This is a believable and easy message to convey when the original appeal is subtle and highly deniable.*" (Mendelberg 2001: 274)

Something like this happened during the 2008 presidential campaign, when Barack Obama suggested that John McCain and others in the GOP would play up the fact that he (Obama) "doesn't look like all those other presidents on the dollar bills."[34] McCain's camp responded by saying, "We are not going to let anybody paint John McCain, who has fought his entire life for equal rights for everyone, (sic) to be able to be painted as racist."[35] Accusing Obama of accusing McCain of being racist forced the Obama camp to backpedal, claiming that "Barack Obama never called John McCain a racist," and that Obama was talking about "his status as a young, relative newcomer to Washington politics."[36] This exchange illustrates the complications inherent in calling out a racial appeal: in doing so, one may be accused of claiming that the appealer is a racist, which, because of the significant social costs of being labeled "racist," may effectively stop the issue in its tracks.[37]

This "table-turning" response to calling out racial appeals works only if (i) it is reasonable to infer from that someone made a racial appeal that they are a racist and (ii) that to be a racist requires some kind of intentional or foreseeable racist behavior. The McCain-style reasoning (also instantiated by Paul Ryan in footnote 11, more on which below) proceeds as follows: (a) the accused racist X did not intentionally or foreseeably act in a racist way, so therefore (b) X is not a racist, and therefore (c) X made no such racial appeal. Now, (ii) is most certainly false, but importantly for our purposes is how the label "racist" is applied by ordinary individuals with respect to prevailing norms of racial equality, and I think many people do accept (ii). But, nonetheless, my theory of code words shows why (i) is false. On my theory, uses of a code word may be effective regardless of what the speaker intended and what she foresaw as effects of her speech. Therefore, we must distinguish the question of whether the speaker using the code word did so with certain intentions (which, given (ii), might constitute a racist action on her behalf) from the issue of whether the use of the code word may succeed in priming racial resentment. We can answer the latter affirmatively without taking a stand on the former, and hence without accusing the speaker of racism.

This is all well and good *theoretically*, but since my theory of code words is not common knowledge, what is the practical upshot for calling out coded racial appeals? My recommendation is that in calling out

---

[33] This is similar to the strategy employed by Jesse Jackson during the Bush 1988 campaign's use of "Willie Horton," which may have had some effect in neutralizing the power of the ad leading up to the election.
[34] http://www.nbcnews.com/id/25966382/ns/politics-decision_08/t/campaigns-trade-barbs-who-started-it
[35] *Ibid*.
[36] *Ibid*.
[37] That "racist" has become such a stigmatized label has ironically resulted in a narrowing of its believed extension via modus tollens: *racists are evil, and since I am not evil, I must not be a racist*.

the racial appeal, we explicitly focus on the appeal itself rather than the speaker making it. It does not matter whether the speaker intended or foresaw anything racial by her statement---that is not the point. Rather, the point is whether the statement allows hearers to act on racist beliefs without appearing to do so.

To see the strategy in action, go back to the exchange between Congresswoman Barbara Lee and Congressman Paul Ryan which we discussed in footnote 11:

(24) *Barbara Lee*: 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.'

(25) *Paul Ryan*: This isn't a race based comment; it's a breakdown of families, it's rural poverty in rural areas, and talking about where poverty exists---there are no jobs and we have a breakdown of the family. This has nothing to do with race.

By accusing Ryan of *intending* to talk about black people when using the phrase 'inner city,' Lee allowed for Ryan to pivot, retreating to the apparently non-racial issue of the breakdown of families in poor areas. However, if instead Lee made it clear she does not care about what Ryan *meant* but rather only about how Ryan's claims will be interpreted, she leaves him no room to pivot in this way. Here is a possible strategy:

(26) *Lee*: Let's be clear, there is a prevailing stereotype that the inner city is primarily home to poor black people. Therefore, Mr. Ryan's comments about the "inner city" will be interpreted by almost everyone as about black neighborhoods in poor urban areas, whether he intends us to or not.

This version of Lee's rebuttal similarly makes explicit the racial component of Ryan's comments. However, it leaves no room for Ryan to pivot away from race, since it nowhere mentions what he meant or intended by his comment---it merely points out explicitly a consequence many listeners will draw from his statement. The hope of this calling out strategy is that it will make explicit the racial component of the coded appeal (thus inhibiting its effects) without leaving room for the speaker to pivot away from race while leaving the implicit message intact.

Concluding remarks

This paper has been an extended discussion of the use of coded speech in political discourse. I have argued for a very minimal theory---that using code words creates space for deniable norm violations not by involving some hidden (coded) meaning in the words used, but rather by exploiting pre-existing stereotypical beliefs. My hope this is that this minimal theory may be all we need to understand coded speech, and may help in further political projects of resisting their effects in political discourse.[38]

Appendix

Differences in the conditions from White 2007:

Verbal Cues in Food Stamps Frames

African Americans	Inner-city Americans	Poor Americans	Working Americans
House democrats stirred debate today on a proposal that would lead to a major restructuring of the nation's welfare system, with claims that the Republican plan would disproportionately hurt black and Hispanic families.	House democrats stirred debate today on a proposal that would lead to a major restructuring of the nation's welfare system, with claims that the Republican plan would disproportionately hurt inner city families.	House democrats stirred debate today on a proposal that would lead to a major restructuring of the nation's welfare system, with claims that the Republican plan would disproportionately hurt poor families.	House democrats stirred debate today on a proposal that would lead to a major restructuring of the nation's welfare system, with claims that the Republican plan would disproportionately hurt working American families.
"… programs like food stamps and Medicaid represent important safety-nets for many African-American families, keeping them from falling deeper into poverty."	"… programs like food stamps and Medicaid represent important safety-nets for many inner-city families, keeping them from falling deeper into poverty."	"… programs like food stamps and Medicaid represent important safety-nets for many poor families, keeping them from falling deeper into poverty."	"… programs like food stamps and Medicaid represent important safety-nets for many American families, keeping them from falling deeper into poverty."
"… this recent round of budget cuts would have a disastrous impact on black Americans," Brown argued. "Thousands of African-American families could lose food stamp benefits and many more would be left without adequate access to affordable health-care …"	"… this recent round of budget cuts would have a disastrous impact on those who live in America's inner cities," Brown argued. "Thousands of inner-city families could lose food stamp benefits and many more would be left without adequate access to affordable health-care …"	"… this recent round of budget cuts would have a disastrous impact on poor Americans," Brown argued. "Thousands of inner-city families could lose food stamp benefits and many more would be left without adequate access to affordable health-care …"	"… this recent round of budget cuts would have a disastrous impact on American families," Brown argued. "Thousands of working American families could lose food stamp benefits and many more would be left without adequate access to affordable health-care …"

References

Abrams, Richard L, Klinger, Mark R, & Greenwald, Anthony G. 2002. Subliminal words activate semantic categories (not automated motor responses). *Psychonomic Bulletin & Review*, **9**(1), 100–106.

Albertson, Bethany L. 2015. Dog-whistle politics: Multivocal communication and religious appeals. *Political Behavior*, **37**(1), 3–26.

Austin, J. L. 1962. *How to Do Things with Words*. 2nd edn. Harvard University Press.

Banaji, Mahzarin R. 2002. Social Psychology of Stereotypes. *In:* Smelser, N., & Baltes, P. (eds), *International encyclopedia of the social and behavioral sciences*. New York: Pergamon.

Bar, Moshe, & Biederman, Irving. 1998. Subliminal visual priming. *Psychological Science*, **9**(6), 464 468.

Beaver, David, Roberts, Craige, Simons, Mandy, & Tonhauser, Judith. 2009. Presupposition, Conventional Implicature, and Beyond: A Unified Account of Projection. *In:* Klinedinst, Nathan, & Rothschild, Daniel (eds), *Proceedings of New Directions in the theory of Presuppositions*. Toulouse: ESSLLI.

Edsall, Thomas Byrne, & Edsall, Mary D. 1992. *Chain reaction: The impact of race, rights, and taxes on American politics*. WW Norton & Company.

Elga, Adam, & Rayo, Agustín. 2016. *Fragmentation and Information Access*. manuscript.

Elliot, Jeffrey. 1993. The 'Willie' Horton Nobody Knows. *The Nation*, **August 23, 1993**.

von Fintel, Kai. 2004. Would You Believe It? The King of France is Back! (Presuppositions and Truth Value Intuitions). *Pages 315–342 of:* Reimer, Marga, & Bezuidenhout, Anne (eds), *Descriptions and Beyond*. Oxford: Oxford University Press.

von Fintel, Kai. 2008. What is Presupposition Accommodation, Again? *Philosophical Perspectives*, **22**, 137–170.

Frege, Gottlob. 1997 (1918). The Thought. *In:* Beany, Michael (ed), *The Frege Reader*. Oxford: Blackwell.

Gibson, Bryan, & Zielaskowski, Katherine. 2013. Subliminal priming of winning images prompts increased betting in slot machine play. *Journal of Applied Social Psychology*, **43**(1), 106–115.

Greenwald, Anthony G, & Banaji, Mahzarin R. 1995. Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological review*, **102**(1), 4.

Greenwald, Anthony G, McGhee, Debbie E, & Schwartz, Jordan LK. 1998. Measuring individual differences in implicit cognition: the implicit association test. *Journal of personality and social psychology*, **74**(6), 1464.

Grice, Paul. 1957. Meaning. *The philosophical review*, 377–388.

Grice, Paul. 1989. *Studies in the Way of Words*. Cambridge: Harvard University Press.

Haslanger, Sally. 2013. Social Meaning and Philosophical Method. *Eastern APA Presidential Address*.

Hurwitz, Jon, & Peffley, Mark. 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), 99 112.

Hutchings, Vincent L, & Jardina, Ashley E. 2009. Experiments on racial priming in political campaigns. *Annual Review of Political Science*, **12**, 397–402.

Kahan, Dan M. 1997. Social influence, social meaning, and deterrence. *Virginia Law Review*, 349 395. Kinder, Donald R, & Sanders, Lynn M. 1996. *Divided by color: Racial politics and democratic ideals*. University of Chicago Press.

Lessig, Lawrence. 1995. The regulation of social meaning. *The University of Chicago Law Review*, **62**(3), 943–1045.

Levinson, Stephen. 2000. *Presumptive Meanings: The Theory of the Generalized Conversational Implicature*. Cambridge: MIT Press.

Lewis, David. 1979. Scorekeeping in a Language Game. *Journal of Philosophical Logic*, **8**, 339–59.

Lewis, David. 1982. Logic for equivocators. *Noûs*, 431–441.

López, Ian Haney. 2014. *Dog whistle politics: How coded racial appeals have reinvented racism and wrecked the middle class*. Oxford University Press.

Lynch, Michael. 2016. Trump, Truth and the Power of Contradiction. *http://www.nytimes.com/2016/05/08/opinion/sunday/trump-truth-and-the-power-of- contradiction.html*

Mandelbaum, Eric. 2015. Attitude, inference, association: On the propositional structure of implicit bias. *Noûs*.

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

Mendelberg, Tali. 2008. Racial priming revived. *Perspectives on Politics*, **6**(01), 109–123.

Morin, Richard. 1988. Behind the Numbers: Confessions Of a Pollster. *The Wall Street Journal*, **October 16**.

Rawls, John. 2005. *Political liberalism*. New York: Columbia University Press.

Schwitzgebel, Eric. 2010. Acting contrary to our professed beliefs or the gulf between occurrent judgment and dispositional belief. *Pacific Philosophical Quarterly*, **91**(4), 531–553.

Sennet, Adam. 2011. Ambiguity. *Stanford Encyclopedia of Philosophy*.

Stalnaker, Robert. 1991. The problem of logical omniscience, I. *Synthese*, **89**(3), 425–440.

Stalnaker, Robert. 1999. The problem of logical omniscience, II. *In: Context and Content*. Oxford: Oxford University Press.

Stalnaker, Robert. 2002. Common Ground. *Linguistics and Philosophy*, **25**, 701–721.

Stanley, Jason. 2015. *How Propaganda Works*. Princeton, NJ: Princeton University Press.

Stewart, Patrick A. 2008. Subliminals in the 2000 presidential election: Policy implications of applied neuroscience. *Public Integrity*, **10**(3), 215–232.

Trump, Donald. 2011. *Time to get tough: Making America # 1 again*. Regnery Publishing.

Valentino, Nicholas A, Hutchings, Vincent L, & White, Ismail K. 2002. Cues that matter: How political ads prime racial attitudes during campaigns. *American Political Science Review*, **96**(01), 75–90.

Weinberger, Joel, & Westen, Drew. 2008. RATS, we should have used Clinton: Subliminal priming in political campaigns. *Political Psychology*, **29**(5), 631–651.

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**(02), 339–354.

Zwicky, Arnold, & Sadock, Jerrold. 1975. Ambiguity tests and how to fail them. *Syntax and semantics*, **4**(1), 1–36.