DOGWHISTLES AND THE AT-ISSUE/NON-AT-ISSUE DISTINCTION*

Robert Henderson
University of Arizona

Eric McCready
Aoyama Gakuin University

1 Introduction

George Bush's 2003 State of the Union address contains the following line.

(1) Yet there's power—wonder-working power—in the goodness and idealism and faith of the American people.

To most people this sounds like, at worst, a civil-religious banality, but to a certain segment of the population the phrase *wonder-working power* is intimately connected to their conception and worship of Jesus. When someone says (1), they hear (2).

(2) Yet there's power—Christian power—in the goodness and idealism and faith of the American people.

In a 2016 Reddit AMA Green Party presidential candidate Jill Stein was asked about the party's platform vaccines and homeopathy. She said:

(3) By the same token, being "tested" and "reviewed" by agencies tied to big pharma and the chemical industry is also problematic.

Even though Stein said she thought vaccines work, across the internet she was accused of being an anti-vaxxer and pro-woo due to phrases like *big pharma*, which to people familiar with alternative-medicine discourses know is demonized as selling poison for profit. They heard:

*We need to thank Heather Burnett, Judith Degen, and * for extremely helpful dicussion in construction the analysis presented here. We also need to thank the participants at * and members of the 2017 DGfS workshop on secondary content.

(4) By the same token, being "tested" and "reviewed" by agencies tied to big pharma and the chemical industry, who sell unsafe vaccines to make a buck, is also problematic.

On a 2014 radio program, Representative Paul Ryan said the following.

(5) 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.

He was criticized shortly after by fellow Representative Barbara Lee for making a "thinly veiled racial attack". This is because the phrase *inner-city* is code or euphemism for African American neighborhoods (espcially stereotypically racialized views of such neighborhoods). Many people heard Paul Ryan say:

(6) We have got this tailspin of culture, in our African American neighborhoods 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.

All three of these examples illustrate the notion of a *dogwhistle*—that is, language that language that sends one message to an outgroup while at the same time sending a second (often taboo, controversial, or inflammatory) message to an ingroup. Dogwhistle language has been explored quite a bit in political science and political economy (e.g., Calfano and Djupe 2008; Goodin and Saward 2005; Hurwitz and Peffley 2005; Mendelberg 2001), and even in their experimental literatures. For instance, Albertson 2015 shows experimentally that examples like (1) do in fact improve a speaker's appeal to religious voters, while slipping right by unreligious voters, unlike uncoded religious appeals like (2), which are punished by non-religious voters. In contrast, the linguistic literature on dogwhistles is practically non-existent. One exception is Stanley 2015, which provides a substantive semantic / pragmatic proposal, where dogwhistles are Pottsian CIs, contributing an at-issue component for the outgroup audience and a non-at-issue component that potentially only the ingroup is sensitive to.

In this paper we argue against a CI account of dogwhistles and instead propose alternative, purely pragmatic account combining aspects of McCready 2012, Burnett 2016; Burnett 2017, and which we think better accounts for the core their core properties.[1] In broad strokes, we make the novel proposal that dogwhistles come in two types. The first concerns covert signals that the speaker has a certain persona, which we model by extending the *Sociolinguistic Signalling Games* of Burnett 2016; Burnett 2017. The second involves sending a message with an enriched meaning whose recovery is contingent on recognizing the speaker's covertly signalled persona.

2 Against a CI account

Recall that in addition to expressions that bear at-issue content alone, there are expressions like slurs, honorifics, etc., which carry a conventional non-at-issue component as well. For example, a slur like *kraut* would have AI-component "German" and a NAI-component "I hate Germans". In general, terms like *kraut* which carry both AI and NAI components can be referred to as *mixed content bearers*. Stanley 2015 argues that dogwhistle language should be analyzed as mixed content bearers. In particular, a dogwhistle like *welfare* would have AI-component "the SNAP program" and a NAI-component "Blacks are lazy". There are a series of reasons to believe that this is not the case.

Our first argument, which we call the *knowledge argument*, is based on what it takes to plausibly say a speakers knows the meaning of a word. We argue that the requirements for knowing the meaning of dogwhistles seem quite different from those for widely accepted cases of mixed content. Take the case of pejoratives. Can a speaker know what *kraut* means without knowing it is derogatory? It seems not. Conversely, can a speaker know what *welfare* means without knowing this association with Cadillacs, etc. (p. 158-9)? We think the answer is: Yes. The whole idea of a dogwhistle is that the (so-called) NAI component is not accessible to some speakers. Thus, the NAI part must not be part of the conventional meaning.

An immediate objecion to the knowledge argument would be that we are just dealing with different dialects? This argument seems to beg the substantive question, but there are other reasons to think it incorrect. While this view might explain the effect of dogwhistles in mixed company, but does not explain the use of dogwhistles with an in-group. Under a dialect account, dog-whistle language should also be what is used when talking to an in-group because this is just what the words mean for the audience. This doesn't seem right. Dogwhistles, by definition, are not needed when talking to an in-group and can be disposed of, which wouldn't make sense if the subtext of dogwhistle were part of its conventional meaning for the in-

[1] In recent work, Khoo to appear proposes an alterative to a CI account. We find his proposal compelling, and it approaches our own intuitions on the topic, but we believe there is room for improvement, though we cannot address the project in this paper for reasons of space.

group. Ultimately what we'll propose is that we have distinct groups of speakers, but the way they are distinct is not a way characteristic of how 'genuine' dialects work, but rather involves different background knowledge about language *use*.

The second argument against a mixed content account of dogwhistles, what we called the *deniability argument*, gets at the heart of what it means for content to be convetional. The use of dogwhistles is prompted by a desire to 'veil' a bit of content, but still to convey it in some manner. Deniability is essential. If a bit of content is conventional, it's not deniable any longer. This can be seen with pejoratives, which clearly carry conventional NAI content.

(7) A: Angela Merkel is a kraut.
B: What do you have against Germans?
A: #I don't have anything against Germans. Why do you think I might?

Such dialogues are fine with dogwhistles; in the following, there seems to be no entailment that A has the relevant attitude.

(8) A: Eric is on welfare.
B: What do you have against social programs?
A: I don't have anything against social programs. Why do you think I might?

eneralizing, we can identify a dialogue-based test for conventional content: in a dialogue in which participant A says 'X', where $[\![X]\!]$ is a mixed content bearer with AI content Y and NAI content Z and participant B responds with 'It's not cool to say Z', it is incoherent for A to respond 'I didn't say that Z" if Z is conventional content. By this test, dogwhistles of all types can be concluded to not be conventional, and thus a fortiori, not mixed content bearers.

We take the knowledge and deniability arguments to be present a strong challenge to the mixed content bearer view of dogwhistle language. In the following sections we develop an account of dogwhistle language that avoids problems with a conventionalized CI analysis, while accounting for what we take to be its core properties: (i) dogwhistles are not part of conventional content, so speakers are able to avoid (complete) responsibility for what they convey, (ii) dogwhistles are semi-cooperative—that is, they are meant to be under-informative to one segment of the audience, while communicating a particular message to another, and (iii) while deniable, dogwhistles are risky. Being detected using a dogwhistle by the wrong party should cost the speaker in some way.

3 Flavors of dogwhistle

While dogwhistle language is often treated as a uniform phenomenon, we think there are two prototypical cases (though they smear into one another). This novel empirical observation will structure the rest of the paper because we provide two separate, but connected analyses of the two types of dogwhistles which can be intuitively described as follows.

In a Type 1 dogwhistle, the content sends one message to all audience members, while the whistle transmits the speaker's true identity to a sub-audience. The Stein and Bush cases above probably best fit in this category. Stein's "Big Pharma" just means large, faceless pharmaceutical corporations (parallel to "Big Agriculture", etc.), but she flagged herself a vaccine denier because that phrase is primarily used in vaccine-denial (and alternative medicine) discourse. Similarly, Bush's "wonder-working power" probably doesn't convey some secondary message about the power at hand, but instead just flags him as an evangelical because only they talk like that.

With Type 2 dogwhistles the content sends one message to all audience members, while the whistle sends places an addendum on that message for a sub-audience. The Ryan case above best fits this category. His use of "inner city" conveys to all audiences a geographical location inside cities, but then to a sub-audience, it specifically picks out African American neighborhoods in those cities. Of course, Ryan's utterance will also allow a listener to infer things about Ryan's identity as in Type 1 examples—this is especially true if the whistle is detected.

We take each case in turn, starting from the simpler Type 1 and then expanding into Type 2. The core idea underlying Type 1 dogwhistles is that they can be handled via sociolinguistic persona construction in a Bayesian signalling game. Type 2 dogwhistles will then involve the same kind of signaling games, but with an extra pragmatic enrichment that will follow the detection of particular speaker personae in such a game.

4 Type 1 dogwhistles & sociolinguistic signalling games

In recent work, Burnett 2016; Burnett 2017 pioneers the use of Bayesian signaling games to model identity construction through sociolinguistic variation. We take Type 1 dogwhistles to be only slightly more complex verions of sociolinguistic identity construction through variation of the kind discussed in Burnett's work.

Social Meaning Games (SMG) in the style of Burnett 2016; Burnett 2017 have a simplified architecture (which we modify / elaborate further below). First, there are two players, a speaker S and listener L. As with all games, players have actions that they can perform, which can potentially depend on and interact with the actions of other players. In SMG, the actions are as follows: (i) The speaker chooses a persona p from the space of personae P, (ii) Based on their persona, the speaker chooses a message $m \in M$ to send to the listener, and (iii) based on the message, the listener chooses a response $r \in R$, which in the simplest case we can identify with selecting an element of P—i.e., identifying the speaker's persona. Finally, we have utilitiy functions for players U_S/U_R—functions from $P \times M \times R$ to $\mathbb{R}$, which represents payoffs for every possible combination of actions. In the kinds of games we consider, the speaker's utility is maximized by picking a message that sends the most information to the listener about the persona they want them to assign to them. The listener's utility is maximized if they extract the most information they can about a speaker's persona given their message.

We now elaborate on these ingredients and model the behavior of Type 1 dog-whistles. First, we need to specify what we mean by a persona. We assume that there is a space of properties that a person can have, and a persona is a collection of these properties. More formally, the set of personae P is a set of maximally consistent sets of properties. For instance, in the Stein case, the relevant properties might be: ANTI-VAX, PRO-VAX, ANTI-CORPORATE, PRO-CORPORATE. The speaker would be trying, though their words, to construction a persona that is a subset of these properties. For instance, the speaker might want to present as, depending on the audience and their preferencies, the yuppie doctor (i.e., {ANTI-CORPORATE, PRO-VAX}) or the hippie radical (i.e., {ANTI-CORPORATE, ANTI-VAX}). Of course, some collections of properties of simply incoherent, like being both pro- and anti-vax. Taking this in to account, as well as the fact that one's corporate stance is, in principle, separable from one's stance on science, we have the following persona-space based on the four properties we identified.

(9) {{ANTI-VAX, ANTI-CORPORATE},
{ANTI-VAX, PRO-CORPORATE},
{PRO-VAX, ANTI-CORPORATE},
{PRO-VAX, PRO-CORPORATE}}

Our SGMs will then revolve around a speaker picking some persona from a given persona-space P, as well as a message that will signal the persona selected. The hearer will in select from P a persona for the speaker based on the message they received.

Implicit in the information description of SMGs is the notion that messages are linked with personae in some way. We make this explicit by treating messages as having two kinds of meanings. To begin, messages $m \in M$ have their normal denotational meaning $[\![m]\!]$. [2] Additionally, messages have *social meaning*, which we identify directly with the personae they signal. That is, given a persona-space P, the social meaning of a message m is just an element of P, namely $[m] \in P$. While a message may denote a particular persona, we recognize that persona construction is more complex and may involve integrating information from multiple messages. That is, just as we may talk of two propositions being consistent or intersecting them to combine their information, we want to be able to talk about messages having consistent social meaning or reason about how two messages together signal a more specific persona. For this reason, we will often have recourse to reference the personae that are consistent with a message message m, which we write $c(m)$ and define as follows.

(10) $c(m) := \{n \in M | m \cap n \neq \emptyset\}$

To give a concrete example in the persona-space discussed above, if the expression *big pharma* has the social meaning in (11), example (12) gives its set of consistent personae. That is, using *Big Pharma* is consistent with any persona that is not $\{$PRO-VAX, PRO-CORPORATE$\}$.

(11) $[$*Big Pharma*$] = \{$ANTI-VAX, ANTI-CORPORATE$\}$

(12) $c($*Big Pharma*$) = \{\{$ANTI-VAX, ANTI-CORPORATE$\}$,
$\{$ANTI-VAX, PRO-CORPORATE$\}$,
$\{$PRO-VAX, ANTI-CORPORATE$\}\}$

Now that we have defined personae and persona-spaces, as well as how they are linked to messages, the fully elaborated action structure for SMGs is as follows. The speaker begins by picking a persona and a message—e.g., $\langle\{$ANTI-VAX, ANTI-CORPORATE$\}$, *Big Pharma*$\rangle$. Then, the listener identifies the speaker's persona based on their message from P in (9), where we know that the social meaning of *Big Pharma* rules out the persona $\{$PRO-VAX, PRO-CORPORATE$\}$.

Given this action structure, we need to define utility functions for both speaker and listener, which should be, as usual with signaling games and their variants, understood as functions from speaker and listener choices to the reals. We can do this via conditionalization, whose general form is given in (13).

(13) $Pr(p|m) = \frac{Pr(p\&m)}{Pr(m)}$
"The probability of persona p given message m"

Burnett 2016; Burnett 2017 specifies how to compute the joint probability of a message and the joint probability of a message and persona; here $Pr(p)$ is given

[2] We won't have much to say about the denotational meaning of messages, though

by the listener's prior beliefs about the speaker's persona which is a probability distribution over P. We will modify this approach in the following for reasons to be discussed shortly.

(14) $Pr(p|m) = \frac{Pr(\{p\} \cap c(m))}{Pr(c(m))}$

Because messages are identified with their social meaning (e.g, [*Big Pharma*] $=$ $\{$ANTI-VAX, ANTI-CORPORATE$\}$), the probability of $p\&m$ when they are consistent is just the probability of the persona—i.e., $Pr(p)Pr(m)$ is the probability of all the personae consistent with m, namely $Pr(c(m))$.

Here is where we have to begin to diverge from Burnett 2016; Burnett 2017. The reason is that we want the dogwhistle effect to arise from listeners being unaware (or uncertain) about the closeness of the connection between some bit of language an a persona. Otherwise stated, we want listeners to have beliefs about a speaker's persona, but also beliefs about how personae and messages are connected. Formally implemented, this amounts to letting listeners have priors over P, but also beliefs about $P(m|p)$—namely how closely messages are linked to particular personae.[3]

Given the above, we can now update a listener's belief about the speaker's persona given their message using Bayes' theorem. Note that the probability of the message can be directly computed.

(15) $Pr(p|m) = \frac{Pr(p)Pr(m|p)}{Pr(m)}$

(16) $Pr(m) = \sum_{p \in P} Pr(m|p)Pr(p)$

While we need this complication for Type 1 dogwhistles, this is a trivial extension of Burnett 2016; Burnett 2017. Burnett's analysis is recovered by just assuming that the likelihood (i.e., $Pr(m|p)$) is 1 whenever m and p are consistent. One way to think about this is that in Burnett's system, the probability of hearing a p-consistent message with p is just the probability of p because messages and personae are perfectly coupled and the mapping between them is in a sense direct. But, for the dogwhistle case, what we allow is for messages to more strongly or weakly signal types that they are consistent with.

The final ingredient we need to provide utility functions is some way to encode the fact that speakers don't just *report* their personae, but construct them in concert with their listeners.[4] Speakers want to present themselves in a certain way. Speak-

[3] The term "belief" might be a bit too strong, to the degree that one takes the content of belief to be available via introspection. Listeners will have formulated, from their experience, some idea about the particular social messages different kinds of people send and at what frequency. This will obviously differ across listeners, but need not be consciously accessible in the general case.

[4] Burnett discusses this in terms of the implementation of a Butlerian view on the construction

ers will also be sensitive to whether listeners will approve of that persona or not. In the adversarial contexts in which dogwhistles come into their own, a speaker might have to juggle presenting a safe persona with a persona they might prefer to present (or prefer to present to another audience that might be listening)—this is when dogwhistle language become useful. Along these lines, we follow Burnett 2017; Yoon et al. 2016 in assuming that the utility calculation takes into account the message's social value, which is given by two functions. First, the speaker has a function ν_S that assigns a positive real number to each persona representing their preferences. Second, the listener has a function ν_L that assigns a real number (positive or negative) to each persona representing their (dis)approval.

We can now calculate the speaker's utility, though we diverge again from Burnett 2016; Burnett 2017. In that work utilities are computed over persona-message pairs, which allows for reasoning about what persona would be useful to convey. We instead focus on what message should be sent given the particular persona structure and how personae might be received. Thus, we consider a generalized formulation which calculates the utility for the message itself, without considering the particular persona it is intended to convey.

Here, the utility is dependent on the affective values of the range of personae consistent with the message, dependent on the likelihood that the particular persona is recovered given the message, as follows:

(17) $$U_S^{Soc}(m, L) = \sum_{p \in [m]} ln(Pr(p|m)) + \nu_S(p)Pr(p|m) + \nu_L(p)Pr(p|m)$$

When only one listener is addressed, dogwhistles reduce to ordinary social meaning; the speaker should choose a signal which maximizes U_S^{Soc}.

Dogwhistles come into their own when speakers address groups of individuals with mixed preference over personae, different priors for the speaker's persona, and different experiences about the likelihood of a persona given a message.The simplest way to assign utilities to the group case is to sum over all listeners; we will assume this metric in the following.

(18) $$U_S^{Soc}(m, G) = \sum_{L \in G} U_S^{Soc}(m, L)$$

Note, though, that we think the method of simple summation should only be taken as a starting point. There are probably cases in which this way of calcu-

of social roles (e.g. **?**), according to which they are dynamically built by agents, so that there is an interplay between society and agents in terms of the range of roles made available, which can be expanded or altered by the behavior of agents (or, at least, this is the way in which we understand this work). We are not convinced that the current model properly models this aspect of social role construction; in the main text we take a more conservative stance, on which the construction happens in concert between speakers and hearers, but drawing on an existing set of available roles.

lating utilities overridden—e.g., if one particular powerful person in the audience is known to have a highly negative affective value for a particular persona which she is likely to recover. Also, in the case of a particularly pernicious persona (i.e. one for which ν_L yields an extremely low value), the possibility of later penalty may preclude the use of the dogwhistle in the first place despite present advantage. Modeling this requires a move to a repeated game setting (cf. McCready 2015) and we will leave its analysis for later work.

Given this, what should speakers do? Note that the speaker's utility depends only on three factors: the informativity of sending the message given the various persona in play, the speaker's value for particular personae, and the speaker's beliefs about the how the listener(s) will (dis)approve of particular personae. This means we can reason about the speaker's behavior without discussing the listener's utility; we require only the aspect of hearer utility that plays into the speaker preferences. We plan to discuss the listener's optimal behavior in future work. In the meantime, though, we turn to a case study of Type 1 dogwhistles from the speaker's perspective.

5 A case study of Type 1 dogwhistles

Jill Stein is in a predicament.

> She has just been asked about vaccines. She knows her base is basically all anti-corporate, but she also knows her base contains a passionate anti-vaxx minority that hold a position others in her party don't like. She knows that this her anti-corporate bona fides are solid, but the question wouldn't be coming up unless there was some uncertainty about her stance on vaccines. She realizes, though, this is the perfect occasion for a dogwhistle. Her audience has only three types of listeners—the passionate anti-vaxxer, the clueless pro-vaxxer, and the knowledgeable pro-vaxxer—and she can satisfy most everyone while maintaining plausible deniability if her strategy is discovered.

We assume that Stein is choosing between messages whose social meanings always mark her has anti-corporate, but mark her as either pro- or anti-vaccine.

Social meanings	Consistent Personae
big pharma	{ANTI-VAXX, ANTI-CORPORATE}
	{ANTI-VAXX, PRO-CORPORATE}
	{PRO-VAXX, ANTI-CORPORATE}
corporate scientists	{PRO-VAXX, ANTI-CORPORATE}
	{ANTI-VAXX, ANTI-CORPORATE}
	{PRO-VAXX, PRO-CORPORATE}

We further assume that Stein takes all listeners to have the following prior about her persona. That is, they believe that she is probably anti-corporate, but it is equally probable that she is pro- or anti-vaxx (which is why the question is being asked). We represent that with the following priors over personae. The particular numbers do no matter so much as the order. Listeners have a larger and equal degree of believe that she is pro- or anti-vax. They have some smaller degree of belief that she is anti-vaxx, but somehow pro-corporate. Finally, we take listeners assign a very low, but still non-zero chance that Stein is both pro-corporate and anti-vaxx.

Personae	Priors
{PRO-VAXX, PRO-CORPORATE}	.05
{PRO-VAXX, ANTI-CORPORATE}	.40
{ANTI-VAXX, PRO-CORPORATE}	.15
{ANTI-VAXX, ANTI-CORPORATE}	.40

She also supposes her audience is polarized on this issue, but there is structure to this polarization. Often constituencies are composed of highly-motivated, warring subconstituencies with a larger center with opinions, but are somewhat less invested. Along these lines, we assume the anti-vaxxers clearly care a lot about the issue, and the savvy pro-vaxxers, as demonstrated by their knowledge of anti-vaxx discoursem care a lot about Stein's stance. If she is detected as liking vaccinations at all, the anti-vaxxers will be angry and savvy pro-vaxxers will love her, and vice versa. We see the following two figures that anti-vaxxers and savvy pro-vaxxers are mirror images of each other.

$\nu_{\mathbf{L}}(\mathbf{p})$ **for Anti-Vaxxers**

Personae	Values
{PRO-VAXX, PRO-CORPORATE}	-100
{PRO-VAXX, ANTI-CORPORATE}	-100
{ANTI-VAXX, PRO-CORPORATE}	100
{ANTI-VAXX, ANTI-CORPORATE}	100

$\nu_{\mathbf{L}}(\mathbf{p})$ **for Savvy Pro-Vaxxers**

Personae	Values
{PRO-VAXX, PRO-CORPORATE}	100
{PRO-VAXX, ANTI-CORPORATE}	100
{ANTI-VAXX, PRO-CORPORATE}	-100
{ANTI-VAXX, ANTI-CORPORATE}	-100

An unsavy pro-vaxxer has more attenuated belief. We assume that if they discover Stein to be anti-vaxx, they will highly object. That said, the vaxx war is not something they are highly invested in. If Stein is detected to be pro-vaxx, they are happy, but it's consider a kind of default position and so not as big a deal as for the savvy pro-vaxxers.

$\nu_{\mathbf{L}}(\mathbf{p})$ **for Unsavvy Pro-Vaxxers**

Personae	Values
{PRO-VAXX, PRO-CORPORATE}	75
{PRO-VAXX, ANTI-CORPORATE}	75
{ANTI-VAXX, PRO-CORPORATE}	-100
{ANTI-VAXX, ANTI-CORPORATE}	-100

While the audience cares a lot about Stein's persona, we assume that Stein completely accomodating to her audience. She has no preferences among personae, and only wants to maxize her audience's reception of her.

$\nu_{\mathbf{S}}(\mathbf{p})$

Personae	Values
{PRO-VAXX, PRO-CORPORATE}	0
{PRO-VAXX, ANTI-CORPORATE}	0
{ANTI-VAXX, PRO-CORPORATE}	0
{ANTI-VAXX, ANTI-CORPORATE}	0

This assumption is probably most accurate for political discourses where the speaker wants above all to have the listener approve of their persona. That said, note that adding a speaker preference will only cause Stein to either dog-whistle in more risky situations (where the audience is perhaps not balanced correctly) or refrain from dog-whistling when it would otherwise be safe to do so. This means we can safely ignore it to keep the examples less complicated.

Finally, Stein believes that listners might not uniformly take certain messages to go with certain persnonae. In particular. She assumes that all anti-vaxxers are savvy about the phrase “Big Pharma” and it’s place in anti-vaxx discourses, but pro-vaxxers either know about Big Pharma or not. Second, all speakers realize that a phrase like "corporate scientists" is pro-vaxx, in virtue of mentioning scientists, but anti-corporate (in virtue of tying those scientsts to corporate interests). Note that we assume anti-vaxxers and savy pro-vaxxers have she same probability structure below—this actaully attenuates the utility of a dogwhistle. The less an outgroup is aware ingroup messaging, the more useful it will be to dogwhistle. We break down for each group the prior probabilities of the message “Big Pharma” given the various personae at issue as follows.

Anti-Vaxx and Savvy Pro-Vaxx likelihoods for “Big Pharma”
While it is possible (consistent with the social meaning of the phrase), that a speaker might use “Big Pharma” to signal they are just anti-corporate, these listeners know this is phrasing used by their anti-vaxx/anti-corporate allies. Also note this phrasing is inconsistent with a pro-vaxx and pro-corporate persona, which we assume speakers know based on knowing the social meaning of the phrase. This is why the conditional probability is zero.

Personae	$Pr(m\|p)$
{PRO-VAXX, PRO-CORPORATE}	0
{PRO-VAXX, ANTI-CORPORATE}	.1
{ANTI-VAXX, PRO-CORPORATE}	.1
{ANTI-VAXX, ANTI-CORPORATE}	.8

Unsavvy Pro-Vaxx likelihoods for “Big Pharma”
Listeners not aware of anti-vaxx discourse consider this phrase to be consistent with an anti-vaxx persona, but is taken to be primarily an anti-corporate phrase. That is, these listeners don’t see the tight connection between “Big Pharma” and anti-vaxx personae. This is why we call them unsavvy.

Personae	$Pr(m\|p)$
{PRO-VAXX, PRO-CORPORATE}	0
{PRO-VAXX, ANTI-CORPORATE}	.7
{ANTI-VAXX, PRO-CORPORATE}	.1
{ANTI-VAXX, ANTI-CORPORATE}	.15

Everyone’s likelihoods for “Corporate Scientists”
Note that this phrase all but rules out being anti-vaxx, but leans anti-corporate.

Personae	$Pr(m\|p)$
{PRO-VAXX, PRO-CORPORATE}	.6
{PRO-VAXX, ANTI-CORPORATE}	.8
{ANTI-VAXX, PRO-CORPORATE}	.1
{ANTI-VAXX, ANTI-CORPORATE}	.1

We now have utilities of the two messages for differnet kinds of listeners based on the formula in (20).

The Savvy Pro-Vaxxer
The Savvy Pro-Vaxxer has a large negative for Big Pharma, which is the effect of detecting that the phrase "Big Pharma" signals anti-vaxx personae, which they disapprove of.

Message	Utility
Big Pharma	-84
Corporate Scientists	64

The Unsavvy Pro-Vaxxer
In comparison, the Savvy Pro-Vaxxer has a much higher utility for "Big Pharma". This is the dog-whistle effect because the Unsavvy Pro-Vaxxer misses the dogwhistle—namely that "Big Pharma" highly codes for the anti-vaxx persona (the residual negative relative to the message "Corporate Scientists" is from the fact that it doesn't rule out anti-vaxx personas like other phrasing. That is, it is more cagey though doesn't implicate the speaker as an anti-vaxxer to these listeners).

Message	Utility
Big Pharma	32
Corporate Scientists	42

The Anti-Vaxxer
Finally, the Anti-Vaxxer shows the opposite pattern from the Savvy Pro-Vaxxer. This is because they also hear the dogwhistle, but endorse its message.

Message	Utility
Big Pharma	73
Corporate Scientists	-81

In comparing these utilities, we already see the Type 1 dogwhistle effect, namely a message's utility can be greatly increased when listeners fail to realize how tightly it's correlated with a persona they disapprove of. For us, the effect is due to that the fact that a listener's (dis)approaval of a persona affects the utility of a message in proportion to probability they assign that persona given the message. If some listeners are unaware that a message tightly signals a persona, their reaction to that persona can be discounted relative to other listeners that are aware (and may have opposing reaction).

While we already have an analysis of what makes a Type 1 dogwhistle a dogwhistle, the model also makes predictions about when it is optimal to deploy such language. In particular, it makes predictions about audience structure. If we sum message utilities over each listener in a audience, the optimal message will depend on the proportion of different types of listeners (the speaker thinks) are in the audience. In general, given n kinds of listeners, it will be optimal to use a dogwhistle over a disavowal it the following equality holds—where x_n is the number of listeners of type L^n.

(19)
$$\begin{aligned}&(x_1 * U_S^{Soc}(\textsc{dogwhistle}, L^1)), ..., (x_n * U_S^{Soc}(\textsc{dogwhistle}, L^n)) > \\ &(x_1 * U_S^{Soc}(\textsc{disavowal}, L^1)), ..., (x_n * U_S^{Soc}(\textsc{disavowal}, L^n))\end{aligned}$$

To evaluate thie formula, let's consider our intuitions about the scenarios. We have the following core intuitions:

1. If Stein thinks she is talking to any number of pro-vaxxers, whether or not that person is savvy about anti-vaxx discourse or not, she is best to issue a disavowal.

2. If Stein thinks she is talking to any number of anti-vaxxers, she should obviously not disavow and instead issue the dogwhistle.[5]

3. If Stein is talking in mixed company, there things are more complicated, but the ratio of anti-vaxxers to pro-vaxxers (of both types) will determine whether it's best to dogwhistle.

   (a) If there are too few anti-vaxxers in the mix, she can afford to alienate them, issue a disavowal, and reap the utility of signalling her pro-vaxx stance to a primarily pro-vaxx audience.

[5] Actually, she may want to issue a direct appeal, but we have not modelled a third explitly anti-vaxx message, though we could. In previous experimental work (e.g., Albertson 2015), listeners who would approve of a direct appeal don't seem to prefer it over the dogwhistle, though this probably depends on their listener model, that is, who they think might else be listening.

(b) In this calculating, the Savvy Pro-Vaxxers matter more than the Unsavvy. That is, the lower the ratio of Savvy Pro-Vaxxers to Anti-Vaxxers, the more Pro-Vaxxers we need in total to make it worth her while to issue a disavowal.

Our model captures this dynamic. First, note that because the utilities for “Big Pharma” / “Corporate Scientists” are $-84/64$ and $32/42$ for Pro-Vaxxers of both types, it is just always better to avoid the dogwhistle if we have a uniformly pro-vaxx audience. Second, note that because the utilities for “Big Pharma” / “Corporate Scientists” are $73/-81$ for anti-vaxxers, it is always best to use the former in a pure anti-vaxx crowd. Finally, some calculations using the formula in (19) shows that we capture our third intuition above. In particular, if there are twice as many Anti-Vaxxers as Savvy Pro-Vaxxers, then it is optimal to use the dogwhistle as long as there aren’t more than 16 times as many Unsavvy Pro-Vaxxers as Savvy Pro-Vaxxers. If we increase the ratio to 4 to 1, then the break-even point is around 50 times the as many Unsavvy Pro-Vaxxers as Savvy Pro-Vaxxers. More concretely, based on the numbers above, if Stein speaking to audience of 5200 people, it will be optimal to use the dogwhistle if she thinks 400 are hard core anti-vaxxers, 100 are pro-vaxxers who follow the anti-vaxx literature, and the rest are pro-vaxx, but not savvy about anti-vaxx rhetoric.

We believe this seems pretty reasonable, though it would be interesting to study experimentally speakers’ tolerances for deploying dogwhistles, given audience size / structure—presumably it has a kind of Weberian distribution that is not modeled here. Most importantly for this prelimary work, though, is that we can explain what makes for Type 1 dogwhistles languages, while capturing that fact that deploying a dogwhistle is only optimal when the audience has the appropriate structure.

6 Extending the account to Type 2 dogwhistles

We now have an analysis in place for Type 1 dogwhistles, and we have seen in some detail how it applies to an example. We are now ready to turn to the more complex case of Type 2 dogwhistles, which, as the reader will recall, are those in which the persona recovered by the hearer contributes to an enrichment of the semantic contnt of the utterance. To analyze Type 2 dogwhistles we import the machinery of standard signaling games. Our basic analytic strategy is to use signaling games in which there are signals with two possible meanings, one an enriched version of the other, and then to let recovery of the enriched version be tied to recognition of the relevant persona by interpreting messages as pairs of truth-conditional meanings and social meanings of the form $\langle m, [m] \rangle$, on which payoff conditions are imposed.

For the domain of non-social-meaning communication, we need some additional game components to reflect the information transmission aspect of communication. This area of study, fortunately, is well-developed both in the general study of signaling games and in linguistic pragmatics, so we have a range of tools available which can be straightforwardly adopted. We basically require a way to model truth-tracking or accurate communication; for this, we need a set of states (worlds) W. We will take messages to reflect both the state of the world, as drawn from W, and the persona(e) the speaker aims to commmunicate. Thus, speaker strategies σ are now functions from pairs of states and personae to messages, and listener strategies ρ are functions from messages to such pairs.

With this assumption, we can set a utility function which takes into account both information retrieval and social meaning. Let $\rho(\sigma(p,t)) = (p', t')$. Then speaker utilities can be defined as follows.

(20) $U_S(m, L) = U_S^{Soc}(m, L) + EU(L, Pr)$, where $EU(L, Pr) = \sum_{t \in T} Pr(t) \times U(t, L)$, where $U(t, L) > 0$ if $t = t'$ and else $= 0$ (cf. van Rooij 2008).

The social meaning is as defined before, but now to it is added the quantity $EU(L, Pr)$, which sums the product of the utilities of the signal in each state and its probability, where the utility is positive just in case the combination of speaker and hearer strategies allows recovery of the state in which the signal was sent. According to this definition, the utility of the social meaning previously defined is always received, but if the listener fails to recover the proper truth-conditional meaning, no value is extracted from this aspect of the communication.

A more elaborated version of this function can be given by weighting the two components of the utilities with values δ, γ, giving the following.

(21) $U_S(m, L) = \delta U_S^{Soc}(m, L) + \gamma EU(L, Pr)$.

Here δ indexes the value placed on the social meaning and γ the value of the truth-conditional meaning. Setting $\delta = 0$ gives an Aspberger's style of communication, where social meaning is disregarded; at the other extreme, setting $\gamma = 0$ gives Donald Trump ('post-truth'), where truth is irrelevant and only social meaning matters. For present purposes, we will use an unweighted version, but as the current political scene makes clear, Trumpian communication is in fact a viable strategy which could play a role in the use of dogwhistles.

In general, the above seems correct; indeed, it seems correct for Type 1 dogwhistles, where the communicated social meanings and truth-conditional meanings are (at least conventionally) independent. But more needs to be said for Type 2 meanings. The reason is that, in these cases, proper recovery of the intended (enriched) truth-conditional meaning is partly dependent on identifying the relevant persona. We are inclined to view this as a kind of pragmatic encroachment somewhat parallel to the cases discussed by e.g. Recanati 2003, as in the following example, which on its literal reading is false, but when enriched by the italicized material, becomes a coherent response to a crying child.

(22) You're not going to die (*from that cut*). (mom to child on the playground)

We think the difference is that standard cases like the above are entirely contextually conditioned, while Type 2 dogwhistles seem to be the result of a conventional association: once the persona is identified, the additional meaning becomes apparent to the interpreter.

This means that Type 1 dogwhistles are actually a special case of Type 2 dogwhistles in which the additional meaning provided is null. In general, there seem to be two steps in the kind of interpretation exemplified by the Type 2 case (and, arguably, by Type 1 cases as well if one views them as a degenerate specimens of Type 2). The listener first recovers the speaker's persona on the basis of the utterance, and then uses the result to determine 'what is said'. In the present setting, this amounts to conditionalizing prior probabilities on the social meaning and using the posterior probabilities to recover the truth-conditional meaning. This can be simply modeled by altering the expected utility computation for the truth-conditional part of (20) to reference posterior probabilities, as represented by Pr' in (23):

(23) $U_S(m, L) = U_S^{Soc}(m, L) + EU(L, Pr')$, where $EU(L, Pr') = \sum_{t \in T} Pr'(t) \times U(t, L)$, where $U(t, L) > 0$ if $t = t'$ and else $= 0$.

Let us look at a schematic example of how this interpretation proceeds. Consider the utterance (5), with its Type 2 dogwhistle. This utterance contains the phrase 'inner cities' which, on its dogwhistled interpretation, means 'African American neighborhoods'. Without recognizing Paul Ryan's persona, this interpretation seems to be very difficult to get; but, once the persona is recognized, it is very easy, given knowledge of the relevant signal. We analyze this process of interpretation as follows. First, the dogwhistle-sensitive hearer recognizes the persona employed by Ryan, which is signaled directly by the use of 'inner cities.' This results in conditionalization on this persona, which in turn drastically raises the probability of the meaning 'African American neighborhoods' to a point at which this meaning is optimally selected.

In general, it seems that knowledge about social personae can play a role in recovering intended meanings. We suggest that dogwhistles are an instance in which they are in fact crucial. It seems likely there are other such cases as well; likely domains might be pejoratives and honorification, but we leave exploration for future work. In general, however, it seems possible to view dogwhistles as a special case of a broader phenomenon: the use of information about the speaker to recover her intended meaning in cases of ambiguity, underspecification, or indeed semantic enrichment, as discussed by McCready 2012, who provides an analysis in terms of signaling games. However, this sort of content is in general fully cooperative. The special feature of dogwhistles is their strategic character, but we suspect that they are just one case in a broader continuum of cases in which the rational use of language is utilized or manipulated by speakers for reasons of strategy, efficiency, or style. The deeper exploration of this continuum also must be left for later work.

7 Conclusion

This paper has argued against a CI account of dogwhistles on which they introduce mixed content, distinguished two types of dogwhistle, both of which convey social personae but only one of which has at-issue content which is influenced by the persona recovered, and modeled the two types using an extension and variant of Burnett's social meaning games.

We may now consider the following question: what is the best characterization of dogwhistles within existing domains of not-at-issue meaning? They don't appear to obviously fall into any of the categories of meaning standardly assumed in the literature on semantics and pragmatics. First, since dogwhistles ordinarily convey new information, they don't seem to be adequately characterized as presuppositional; it would be quite odd to view them as checking conditions in the common ground. Second, as we have argued, contra Stanley, CIs are an improper characterization, for the meaning is not fully conventional. Rather, on our analysis, all the action in Type 1 dogwhistles is in the domain of social meaning, which in turn is inferred using information about speech styles and social character, while Type 2 dogwhistles further build on the result of these inferences to alter or enrich at-issue content.

Dogwhistles share with conversational implicatures the property of being cancellable (deniable), but differ from standard views of them in not following from anything but an extremely nonstandard construal of the Gricean maxims **grice75** Of course, they are connected with rational language use, so in this sense they have connections with conversational implicature, but the fact that their interpretation arises from background assumptions about social meaning and how personae are linguistically expressed makes them quite different from ordinary implicatures. They are simultaneously conventional and socially dependent. In this sense, dogwhistles seem to occupy a genuinely new niche in the characterization of not-at-issue meaning.

This area is rich with connections to other areas of semantics and pragmatics, to other aspects of social meaning and sociolinguistics, and to other kinds of political speech. Consequently, there are many directions for further work. Let us close with some immediate next steps for this project. First, we want to further consider different kinds of mixed audiences, examining factors such as the influence of degree of sympathy/antipathy and the relative size of the sympathetic group vs unsympathetic savvy vs unsavvy interpreters to yield a more abstract characterization of the function of dogwhistles, together with experimental confirmation of our interpretative model. Second, we plan to look at how dogwhistles behave, and how speakers use dogwhistles, in the case of more extended and complex interactions; in the current setting, this can be done by introduction of a repeated game model with concomitant potential for modeling punishment behavior and possible influence on use of dogwhistles (because of new risk in further interactions, cf. McCready 2015). In the repeated game setting, also, consideration of post-dogwhistle communication between audience members becomes salient: savvy listeners can make unsavvy ones aware of the dogwhistle, influencing subsequent interactions. There is thus interesting interaction between the repeated game structure and epistemic and dynamic logics. However, for this we also need a better understanding of what is gained (or potentially lost) by further signalling of personae; currently our use of ln predicts a substantial loss of value after initial learning, which needs to be looked at further. Finally, we plan on extending the model to other kinds of enrichment phenomena and cases in which social personae interact with interpretations of truth-conditional content. There is a sense in which dogwhistles are an ubiquitous phenomenon: much communication involves underspecified meanings which can in part be resolved by learning more about the speaker and her intentions. Information about social categories often informs how such underspecification is resolved, but possibly in quite different ways in different contexts; this area is also a rich and complex one ripe for investigation.

References

Albertson, Bethany L (2015). “Dog-whistle politics: Multivocal communication and religious appeals”. In: *Political Behavior* 37.1, pp. 3–26.

Burnett, Heather (2016). “Signalling Games, Sociolinguistic Variation and the Construction of Style”. In: *the 40th Penn Linguistics Colloquium, University of Pennsylvania*.

— (2017). “Sociolinguistic Interaction and Identity Construction: The View from Game-Theoretic Pragmatics”. In: *Linguistics and Philosophy*.

Butler, Judith (1990). *Gender Trouble*. Routledge.

Calfano, Brian Robert and Paul A Djupe (2008). “God talk: Religious cues and electoral support”. In: *Political Research Quarterly*.

Goodin, Robert E and Michael Saward (2005). “Dog whistles and democratic mandates”. In: *The Political Quarterly* 76.4, pp. 471–476.

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

Khoo, Justin (to appear). “Code words in political discourse”. In: *Philosophical Topics*.

McCready, Eric (2012). “Emotive equilibria”. In: *Linguistics and Philosophy* 35.3, pp. 243–283.

— (2015). *Reliability in Pragmatics*. Oxford University Press.

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

Recanati, Francois (2003). *Literal Meaning*. Cambridge University Press.

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

van Rooij, Robert (2008). “Game Theory and Quantity Implicatures”. In: *Journal of Economic Methodology* (15), pp. 261–274.

Yoon, Erica J et al. (2016). “Talking with tact: Polite language as a balance between kindness and informativity”. In: *Proceedings of the 38th Annual Conference of the Cognitive Science Society*. Cognitive Science Society.