

How Dogwhistles Work*

R. Henderson and E. McCready

¹ Department of Linguistics
University of Arizona
rhenderson@email.arizona.edu

² Department of English
Aoyama Gakuin University
mccready@cl.aoyama.ac.jp

Abstract. The paper focuses on the semantics and pragmatics of dogwhistles, namely expressions that send one message to an outgroup while at the same time sending a second (often taboo, controversial, or inflammatory) message to an ingroup. There are three questions that need to be resolved to understand the semantics and pragmatics of the phenomenon at hand: (i) What kind of meaning is dogwhistle content—implicature, conventional implicature, etc; (ii) how do (some but not all) hearers recover the dogwhistle content, and (iii) how do expressions become endowed with dogwhistle content? These three questions are interrelated, but previous analyses have emphasized answers to a subset of these questions in ways that provide unsatisfactory answers to the others. The goal for this paper is to take stock of existing accounts, while showing a way forward that reconciles their differences.

1 Introduction

Dogwhistles can be defined as terms that send one message to an outgroup while at the same time sending a second (often taboo, controversial, or inflammatory) message to an ingroup. They are commonly deployed in political contexts to express opinions that politicians calculate will be unpalatable to some segment of their audience but which still yield some advantage when communicated to some other segment. Consider a recent example. On a 2014 radio program Representative Paul Ryan made the following statement, which was criticized shortly after by fellow Representative Barbara Lee as a “thinly veiled racial attack”.

- (1) 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.

* The authors would like to acknowledge the support of JSPS Kiban C Grant #16K02640, which partially supported this research, and to thank Heather Burnett and Judith Degen for extremely helpful discussion during the construction of the analysis presented here, as well as the participants at the 2017 Szklarska Poreba Workshop on the Roots of Pragmasemantics, the 2017 DGfS workshop on secondary content, LENLS 14 and the audience at Nagoya Gakuin University, especially Yu Izumi.

The “thin veil” refers to the phrase *inner-city*, which is code or euphemism for African American neighborhoods (especially stereotypically racialized views of such neighborhoods). Those aware of the dogwhistle heard Ryan promulgate the pernicious racial stereotype that African Americans are lazy. At the same time, Ryan maintained an amount of plausible deniability because this content was not explicit but only referenced a vague geographical location.

We see three questions that need to be resolved to understand the semantics and pragmatics of the phenomenon at hand: (i) What kind of meaning is dogwhistle content—implicature, conventional implicature, etc; (ii) how do (some but not all) hearers recover the dogwhistle content, and (iii) how do expressions become endowed with dogwhistle content? These three questions are interrelated, but previous analyses have emphasized answers to a subset of these questions in ways that provide unsatisfactory answers to the others. The goal for this paper is to take stock of existing accounts, while showing a way forward that reconciles their differences.

We will start by examining these views with an eye to their strengths and weaknesses before pushing forward with our own analysis, which builds on [5]. We take the canonical example above as our test case in evaluating previous proposals and developing our own. As we will show, previous analyses all have their own strengths, but each either fails to address key aspects of the phenomenon or mischaracterizes the content of dogwhistles in problematic ways. The final proposal is that dogwhistles are neither Pottsian CIs as in the work of [16], nor hearer-centered, norm-violating inferences as in the account of [6], but are instead expressions that activate relevant normality conditionals in ingroup speakers but not in outgroup speakers; this can be viewed as a kind of audience-specific invited inference dependent on background knowledge in a way specific to dogwhistles. This analysis improves on the work of Potts and Khoo and fills a lacuna in the game-theoretic analysis of Henderson and McCready, ultimately yielding a substantial advance in our theoretical understanding of dogwhistled speech.

2 Views

There are three views on dogwhistles in the semantics and pragmatics literature. We consider them one by one and compare their perspectives on how dogwhistles work. What we will see is that while it is impossible to maintain a CI view of dogwhistles given the empirical facts, neither the inferentialist view of [6] and the game-theoretic view of [5] manage to account for certain aspects of how dogwhistles work. What we will propose in the light of these observations is a theory that blends aspects of the latter two theories, improving on both.

The CI View

Stanley (2015) is the first published view of dogwhistle speech in semantics and pragmatics. According to this view, dogwhistle terms introduce conventional implicatures: for instance, a term like *welfare*, which is associated with negative

views of poor black people in the US, in addition to its literal meaning, carries the conventional implicature that the speaker dislikes poor blacks. This makes it a mixed content bearer in the sense of [10], though Stanley himself doesn't use this term. However, the dogwhistled content has few to no properties of conventional implicatures, as pointed out by [5] and [6].

Most prominently, if a bit of content is conventional, as the not-at-issue content we see with mixed-content bearers is, it's not deniable any longer. This can be seen with pejoratives, which clearly carry conventional not-at-issue content, roughly that the speaker dislikes the group which is the target of the pejorative expression. It is very odd for a speaker to use a pejorative and then deny the expression of a negative attitude, as in the following dialogue: this is expected given that expressive items and conventional implicatures commit the speaker to the expressed not-at-issue content, as claimed by [12, 13], and given that the not-at-issue content of pejoratives falls into one of these two classes.

- (2) 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. This is unexpected if dogwhistled content is indeed the result of dogwhistles being mixed content bearers.

- (3) A: Donald 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?

By this test, dogwhistles of all types can be concluded to not be conventional, and thus, a fortiori, not mixed content bearers.

The inferentialist view

Another proposal on the market is the inferentialist view of [6], which we are quite sympathetic with in many respects and which addresses aspects of dogwhistles that [5] do not. Khoo's idea is that dogwhistles induce certain kinds of inferences: namely, those which the existing beliefs of interpreters coupled with the information provided by the dogwhistle combine to yield. Schematically, if the speaker claims that x is C and the interpreter believes that C 's are R 's, then the interpreter will conclude that x is R ; it's this kind of inference that Khoo thinks that dogwhistles license. The key point is that if the interpreter lacks the belief that C s are R s, the relevant inference won't arise: this is the way in which Khoo explains the difference in interpretation between ingroup and outgroup speakers. In the example we have focused on, suppose that the interpreter believes that *inner-city* neighborhoods are African American neighborhoods. Then the speaker saying that people who live in inner-city neighborhoods lack a culture of work licenses the inference that people who live in African American neighborhoods lack a culture of work. This is a kind of invited inference account

which relies on the (at-issue) content of the dogwhistle itself and the background beliefs interpreters have which license a constellation of inferences about things related to that content.

This kind of account gets around the problems of treating dogwhistles as CIs. Most importantly, the dogwhistle effect is not conventionalized, but is instead entirely listener-based, which preserves the speaker's deniability. This is the critical fact that CI accounts miss. At the same time, an account that is based entirely on the extensional content of the dogwhistle and the listener's background beliefs is too weak. As Khoo himself notes, the account predicts that any two coextensive terms should induce the same 'dogwhistle inferences,' but they don't: only certain terms do, namely those which can independently be identified as signaling certain aspects of speaker identity in the dogwhistley manner. We thus seem to require a theory in which the dogwhistle inferences are tied to specific linguistic expressions, but not a part of their conventionalized semantic meaning, as with a CI. This is a tricky middle way to find.

Khoo's solution is to appeal to work of [4] on belief fragmentation for a fix: the idea is that thinking of Xs in one way may not deliver the same inferences as thinking of them in a different way, so even coextensive terms may not give the same inferences. Indeed, he indicates (in his footnote 19) that a metalinguistic theory of these words is probably needed (though he makes no attempt to spell one out); 'beliefs about the code words themselves may be relevant.' We agree with this suggestion, though no detail is provided; we think it's precisely the use of the dogwhistle *qua* particularized expression that has to be taken into account when trying to compute what meaning is transmitted and what the likely intentions of the speaker are. However, we want to go further; we suggest that the need to look more closely at the linguistic expressions invalidates the inferentialist theory with its focus on *content*, and requires us to move to a view which induces the inferences arising from dogwhistles on the basis of the forms of the messages themselves, as [5] also emphasize.

The game-theoretic view

[5] provide an account of dogwhistles in a game-theoretic setting. In particular, we build off of pathbreaking work by [1, 2] on what she calls *Sociolinguistic Signalling Games* (SSGs). The core idea behind SSGs is that, in communication, speakers attempt to construct a sociolinguistic persona that listeners try to recover (as in Third Wave sociolinguistics, e.g. [3]). This process is mediated by linguistic expressions, which not only have a semantic meaning, but also a social meaning—namely those personae that the expression is consistent with. Finally, we assume that both speakers and listeners assign values to personae. In the speaker's case, the value is based on rankings on personae: the higher the ranking, the more the speaker wishes to be perceived as having that persona; in the listener's case, it is based on a ranking of personae that they (dis)approve of.

Against this backdrop, dogwhistle language arises under three conditions. First, a linguistic expression becomes associated with a particular persona. This

sort of association is not so surprising. Certain groups of people speak a certain way, and any variation, including lexical choices can signal group membership, a familiar point in sociolinguistic theory (e.g., [7]). Second, there is differential awareness in the population about how strongly that expression signals a particular persona. This is also not surprising. People not aware of a group's culture will not be aware of how they use language (e.g., [15]). Finally, there is a difference in the population in how individuals value that persona. If some individuals value a particular persona highly, but others strongly disapprove of it, there will be an incentive for speakers to signal their adoption of that persona only to certain groups. Under these conditions, it may become possible to use a linguistic expression to signal your persona to a subaudience, while hiding your persona from a large subaudience that would disapprove of that persona. In the game-theoretic perspective, this becomes a utility-maximizing strategy.

Note that in this discussion of persona, we have not talked about communication of 'genuine' linguistic content (ie. plain vanilla semantic content), such as the kind of inference observed in the move from "inner city" to "African American neighborhoods" in the example we have been considering here. [5] show that there are actually two kinds of dogwhistle. Only the first involves exclusively the transmission of speaker personae. With the second kind (called Type 2), the content sends one message to all audience members, while the whistle enables the placement of an addendum on that message for a sub-audience which has a truth-conditional impact, something in the manner of the pragmatic enrichment of [14]. We argue that recovering this message is based on a listener recovering a particular persona for the speaker. The Ryan case above best fits this category. His use of "inner city" may convey to a subaudience that he has a particular persona, and so in virtue of that persona, when he says "inner city" he is referring to African American neighborhoods in those cities.

This kind of account is able to avoid the problems of the CI account. The reason is that speakers, in general, are able to deny that they have a particular persona in virtue of that fact that expressions often only loosely signal particular personas. In the case of Type 2 dogwhistles, denying the persona amounts to denying enriched content sent by way of that persona, e.g., the enriched meaning "African American neighborhood" from the expression "inner city".

What the account in [5] cannot do is explain how "inner city" is related to "African American neighborhood". Assuming that the connection between these two pieces of content is available, our game-theoretic account can provide an account of when the inference from one to the other will arise and that it is deniable, but unlike the account in [6], the relation between the dogwhistle expression and the dogwhistled content is opaque. This is an unfortunate feature of the game-theoretic account, which makes it incomplete as a full account of the pragmatics of dogwhistled expressions. We aim to rectify this situation in the remainder of the paper.

3 A mixed view: Defaults, backgrounds, and form.

From the three views described we can extract a number of considerations relevant to a full account. From the CI account and its problems it becomes clear that dogwhistled content is pragmatic, but not fully conventionalized as a ‘proper’ nondefeasible part of meaning. From the inferentialist account we find a connection between prior beliefs and the content speakers can recover from dogwhistled communication, and also learn that the relevant inferences have to be conditioned on the form of the dogwhistle rather than just its extensional content. From the signaling game view, we see that considerations of utility maximization in particular communicative settings can explain the use of dogwhistles, but not how they arise initially. Our aim now is to bring these insights together into a single unified view.

Our ideal, then, is a theory of dogwhistles which (i) has a metalinguistic character, (ii) makes use of background, default information about how speakers use language, and (iii) can be reconciled with a game-theoretic account of the deployment and recovery of dogwhistled content. Fortunately we have the foundation of such a theory available off-the-shelf in [11], who uses just these components to analyze the way emotive underspecification is resolved in emotive adjectives. We will make use of this theory to underpin our refitting of the inferentialist view to solve its problems with extensionally equivalent expressions and integrate it with our game-theoretic model.

[11] considers cases of underspecified emotive content, which include adjectives like *damn* or *fucking*, particle exclamations like *Man!* and ordinary exclamatives like *What a hotel!*. All these expressions can be interpreted either as positive or negative in the right context.

- (4) What a hotel!
- a. We enter the hotel room: ocean view, palatial space, spotless white coverlet, bottle of champagne, etc. \rightsquigarrow **positive interpretation**
 - b. The hotel is a complete dump, roaches, springs coming out of the bed, plus the window doesn’t open and the AC is broken \rightsquigarrow **negative interpretation**

How can a hearer settle on a positive or negative interpretation? And how can a speaker navigate the potential of hearer misunderstanding of her intention? As these questions make clear, this case is a kind of toy version of the general problem of interpretation recovery in language. The strategy proposed by [11] for this is to condition the interpretation of underspecified emotives on hearer guesses (on the basis of her existing beliefs, modeled using a standard probability function) about speaker emotional states and background knowledge about how people use language with respect to their emotional states. Thus, given that a speaker is using an emotive expression with an underspecified interpretation, that (e.g.) in context (4a) she is in a positive emotional state, and that speakers, when they use underspecified emotives, ordinarily use them in a way that matches their intended interpretation with the speaker’s current emotional state, the hearer

can conclude that the speaker likely means to communicate a positive meaning with the exclamation.

The basic ingredients of the analysis are the use of normality conditionals formulated in a default logic to formalize the way in which particular states of the world associate with emotional states on the one hand and with how language is normally used and interpreted on the other. The first kind of conditional is less relevant to the current setting; it's the metalinguistic aspect of the analysis that makes it useful for overcoming the extensional identity problem. For the case above, we might have conditionals like the following (here, ' $>$ ' is a normality conditional and $\lambda x.Emot(x)$ is a function yielding the emotional state of x , drawn from the set $\{pos, neg\}$):

$$(5) \quad reach_hotel(x) \wedge \exists y[roach(y) \wedge in_room(y)] \wedge \exists z[ac(z) \wedge in_room(z) \wedge broken(z)] > Emot(x) = neg$$

$$(6) \quad reach_hotel(x) \wedge \exists y[champagne(y) \wedge in_room(y)] \wedge \exists z[coverlet(z) \wedge in_room(z) \wedge spotless(z)] > Emot(x) = pos$$

These sorts of world knowledge axioms directly yield the speaker's emotional state, though defeasibly. The use of axioms about metalinguistic content is more general; instead of axioms relating to clearly specified situations like the above, we instead have schema indicating how speaker emotional state ordinarily relates to language use. For example, consider the following axiom from [11], which states that in the absence of defeaters, we can assume that any underspecified emotive expression included in a speaker's utterance should be interpreted in a way conforming with her overall emotional state:

$$(7) \quad (Emot(s^c) = E \wedge Use(s^c, S) \wedge EC \sqsubseteq S) > EC = E$$

Taking the above three axioms together, we are able to arrive at the proper (default) interpretation of the exclamation by chaining the applicable world knowledge axiom with the metalinguistic axiom about interpretation.

Our method for unifying the theories of dogwhistles we have discussed is to import the machinery of defaults about how language is used into our game-theoretic analysis. The result is a theory that blends aspects of [5] and [6], improving on both. The core idea is that the default inferences Khoo uses to arrive at the dogwhistle message should be conditioned on the persona the speaker is aiming to present.

[5] provide a game-theoretic account of how speakers can recover the persona a speaker aims to present (to a subset of the audience) given the choice between distinct, though semantically equivalent expressions. When combined with Khoo's insight about the form of dogwhistle inference, we get an account of how speakers and hearers coordinate on the meaning of dogwhistles. This account is given by the logical formula in (8). According to this view, the speaker's use of a dogwhistle with content C and persona p licenses the inference that the speaker wants the hearer to believe R , given that hearer believes that C entails R in a Khoo-style inference. In (8), $[DW]$ indicates the dogwhistle itself

qua linguistic form; this means that, if the speaker doesn't use the actual dogwhistle, the inference won't follow, so extensionally equivalent expressions won't do the job. In fact, the inference depends on the speaker's persona (because of the clause $Use(s, p, [DW])$), which is connected to the words used, so the extensionally equivalent case is doubly out. Moreover, this account explains why dogwhistles have the enriched meanings they do, which was the aspect of dogwhistle pragmatics left out of [5]. In particular, they get their meanings from the speaker inviting the hearer, based on the speaker's persona, to make an inference they are prone to make based on the content of the dogwhistle.

$$(8) \quad Use(s, p, [DW]) \wedge Bel(h, \forall x[C(x) \rightarrow R(x)]) > Intend(s, Bel(h, R(x)))$$

for persona p and DW with dogwhistled content C .

Let's consider a specific example, that of *inner city*, which has been our test case throughout. This is a Type 2 dogwhistle which, given the recognition of the speaker's persona (cryptoracist), allows enrichment of the content *inner city* to *urban African American neighborhood*. The following axiom states that, given that a speaker s with persona p uses the dogwhistle *inner city*, and given that the hearer believes that inner city neighborhoods are all African American, then normally the speaker intends the inference from his phrasing to this enriched meaning to be made. This captures the proper meaning of the expression, and does so in a way that won't allow substitution of an extensionally equivalent expression such as *neighborhood in an urban area* to induce the relevant inference, for, given the form of the conditional on which the use of the precise expression [*inner_city*] is required, the inference won't be triggered in the absence of either recognition of the speaker's persona or the use of the dogwhistle itself.

$$(9) \quad Use(s, p, [inner_city]) \wedge Bel(h, \forall x[inner_city(x) \rightarrow urban_AA_neighborhood(x)]) > Intend(s, Bel(h, urban_AA_neighborhood(x)))$$

4 Conclusion

In this paper, we have presented three views of dogwhistle language in formal semantics and pragmatics, each flawed in its own way. The first takes them to be CIs; this theory fails due to the lack of conventional quality in dogwhistles. The second is an inferentialist view that aims to explain the kinds of semantic consequences that arise from using dogwhistles, by conditioning them on the background knowledge of those who hear them. This view is a major advance, but fails to explain the fact that dogwhistles don't have the same effects as other, extensionally equivalent, expressions; this is unexpected if only standard kinds of inference are in play, because such inferences are strictly content-based. The third theory conditions dogwhistled interpretations (that affect or enrich truth-conditional content) on the recovery of speaker personae in the sense of Third Wave sociolinguistics; this theory is successful, but makes no attempt to explain

the precise kinds of enrichment that arise with dogwhistles, or why that precise content arises.

The present paper unified the second and third views by making use of metalinguistic default conditionals which condition on the perceived persona of the speaker: since personae are recovered on the basis of the particular expressions used, the choice of expression is important, and extensionally equivalent items won't induce the same effects. This unification is a step forward in the analysis of dogwhistles, and shows further the importance and usefulness of personae in formal pragmatic work.

There are many future directions for this project. Here are two immediate ones. To our knowledge, there is no formal work on dogwhistles outside of English. We are interested to examine dogwhistles in political speech couched in other languages, and, importantly, in non-Western cultures; the results of such investigation should go further to show the (non)uniformity of Gricean considerations of rationality in speech and communication (cf. [9, 8]). An initial test case in this regard might be the Japanese expression *junsui* 'pure', which has a common use in right-wing discourse and dogwhistles views about ethnic purity and Japanese racial homogeneity. Second, we want to look at phenomena which might be viewed as the inverse of dogwhistles in a certain sense: dogwhistles are centered on the recovery of speaker personae, but what about the hearer? Are there expressions which have particular meanings depending on whether the hearer has a particular persona, and lacks them otherwise? We believe so: one such is the phenomenon of *subtweeting*, where a general statement is made which is meant to apply to some specific hearer or set of hearers, possibly just those with a certain property. Thus, hearer self-identification as a target of the subtweet is required for its efficacy. In some cases (perhaps only when the target involves ascription of a property, such as *being a Nazi*), persona identification will be involved; we mean to tease apart some such cases as the next stage in the present project.

References

1. Burnett, H.: Signalling games, sociolinguistic variation and the construction of style. In: the 40th Penn Linguistics Colloquium, University of Pennsylvania (2016)
2. Burnett, H.: Sociolinguistic interaction and identity construction: The view from game-theoretic pragmatics. *Linguistics and Philosophy*. To appear. (2017)
3. Eckert, P.: *Jocks and Burnouts: Social Identity in the High School*. Teachers College Press: New York (1989)
4. Elga, A., Rayo, A.: Fragmentation and information access (1966), manuscript
5. Henderson, R., McCready, E.: Dogwhistles and the at-issue/non-at-issue distinction. In: Gutzmann, D., Turgay, K. (eds.) *Secondary Content: The linguistics of side issues*, pp. 1–21. Brill (to appear)
6. Khoo, J.: Code words in political discourse. *Philosophical Topics* 45(2), 33–64 (2017)
7. Labov, W.: *The social stratification of English in New York City* (1966), center for Applied Linguistics

8. Machery, E.: Expertise and intuitions about reference. *Theoria* 73, 37–54 (2012)
9. Machery, E., Mallon, R., Nichols, S., Stich, S.: Semantics, cross-cultural style. *Cognition* 92, B1–B12 (2004)
10. McCready, E.: Varieties of conventional implicature. *Semantics and Pragmatics* 3, 1–57 (2010)
11. McCready, E.: Emotive equilibria. *Linguistics and Philosophy* 35, 243–283 (2012)
12. Potts, C.: *The Logic of Conventional Implicatures*. Oxford University Press (2005)
13. Potts, C.: The expressive dimension. *Theoretical Linguistics* 33, 165–198 (2007)
14. Recanati, F.: *Literal Meaning*. Cambridge University Press (2003)
15. Rickford, J.R., King, S.: Language and linguistics on trial: Hearing Rachel Jeantel (and other vernacular speakers) in the courtroom and beyond. *Language* 92(4), 948–988 (2016)
16. Stanley, J.: *How Propaganda Works*. Princeton University Press (2015)