# Signaling without Saying

Chapters 2-8

Notes by Robert Henderson
rhenderson@arizona.edu
book with errata fixed: http://bit.ly/45pCyUG
these notes: http://bit.ly/413QaTl

August 6, 2025

## Contents

# Chapter 2: Dogwhistles as a Semantic / Pragmatic Phenomenon

## 1 Introduction

**Chapter Goals**

- This chapter situates dogwhistles within the literature of semantics and pragmatics.
- We will interrogate various previous approaches to dogwhistles, with our eye at this point on the "folk definition" of dogwhistles—i.e.,
- expressions that send one, often taboo message to an in-group, while sending a second innocuous message to an out-group
- We will explore what these messages are, how a speaker sends them, and how they can be received.

**The Central Thesis**

- The main argument is that at its heart, dogwhistling involves *social meaning*.
- Social meaning is a unique, non-propositional category of meaning.
- Despite being non-propositional, pragmatic inferences can still be made over social meaning.
- Typology: We will consider whether dogwhistles are a monolithic category and propose a new, linguistically relevant distinction.

## 2 Previous Accounts

**Part 1: Previous Accounts**

- We will now survey some of the major previous accounts of dogwhistles.
- While understudied compared to phenomena like slurs, a variety of accounts exist.
- The key question that divides these accounts is that of **conventionalization**.

**Parameters of Disagreement**

Previous authors differ on:

- Whether dogwhistles bear some kind of conventional meaning.
- If they do, what the precise content is that is conventionalized.
- If they do not, how to generate the dogwhistle inferences pragmatically.
- Whether these inferences use standard Gricean mechanisms or some other kind of reasoning.

## 2.1 The Conventional Implicature (CI) View

### The Conventional Implicature (CI) View

- Some expressions have both at-issue (AI) and not-at-issue (NAI) content.
- NAI content is often called a conventional implicature (CI).
- Terms with both AI and NAI components are called *mixed content bearers*.

### Example: Slurs as Mixed Content Bearers

- A slur like *kraut* has two meaning components:
  - **AI-component**: "German".
  - **NAI-component**: "I hate Germans".
- The NAI-component is a conventional, not-at-issue part of the word's meaning.

### The CI View of Dogwhistles

- Stanley (2015) argues that dogwhistles should be analyzed as mixed content bearers.
- For example, the dogwhistle *welfare* would have:
  - **AI-component**: "the SNAP program".
  - **NAI-component**: "Black people are lazy".
- There are several reasons to believe this is not the case.

### Argument 1 Against CI: The Knowledge Argument

- The requirements for knowing the meaning of dogwhistles seem different from mixed content bearers like slurs.
- **Question**: Can a speaker know what *kraut* means without knowing it is derogatory?
- **Answer**: It seems not.
- **Question**: Can a speaker know what *welfare* means without knowing its racialized associations?
- **Answer**: Yes. The whole point of a dogwhistle is that the so-called NAI component is not accessible to some speakers.
- Therefore, the NAI part must not be part of the conventionalized truth-conditional meaning.

**Counter-Argument: Dialectal Variation?**

- An objection is that we are just dealing with different dialects.
- On this view, it's not surprising that meaning knowledge varies if the speakers' dialects differ.
- However, this fails to explain the use of dogwhistles even within an in-group.
- A speaker might choose a racial dogwhistle over a racial slur to appear less strident, even when surrounded by fellow racists.
- This strategic choice wouldn't make sense if the dogwhistle's subtext were simply part of its conventional meaning for that dialect.
- A CI account wrongly collapses the distinction between a dogwhistles and, say slurs.

**Argument 2 Against CI: The Deniability Argument**

- This argument gets at the heart of what it means for content to be conventional.
- The use of dogwhistles is prompted by a desire to veil content but still convey it.
- **Deniability is essential**.
- If a bit of content is conventional, it's not deniable any longer.

**Deniability Test: Slurs vs. Dogwhistles**

**Slur Dialogue (Undeniable)**:

**A:** Angela Merkel is a kraut.

**B:** What do you have against Germans?

**A:** # I don't have anything against Germans. Why do you say that?

**Dogwhistle Dialogue (Deniable)**:

**A:** Elin is on welfare.

**B:** What do you have against poor people?

**A:** I don't have anything against poor people. Why do you say that?

**Conclusion on the CI View**

- We can generalize a dialogue-based test for conventional content.
- If B challenges the NAI content Z of an utterance, it is incoherent for A to deny having said Z.
- By this test, dogwhistles are not conventional in this sense.
- The knowledge and deniability arguments present a strong challenge to the mixed content bearer view of dogwhistles.

## 2.2 The Inferentialist View

### The Inferentialist View

- This is the polar opposite of the CI view, proposed by Khoo (2017).
- It's a deflationist position: there is essentially no special linguistic phenomenon at work in dogwhistles.
- The idea is that dogwhistles simply induce certain kinds of inferences.

### How Inferentialism Works

- The effect is generated by the combination of the speaker's utterance and the interpreter's existing beliefs.
- **Schematically**: If a speaker says $x$ is $C$, and the interpreter believes that $C$s are $R$s, the interpreter will conclude $x$ is $R$.
- **Example**: If a speaker talks about people in "inner-city neighborhoods" and the listener believes these are "Black neighborhoods," they will infer the speaker is talking about Black neighborhoods.
- This is a kind of invited inference account.

### Strengths of the Inferentialist View

- It gets around the problems of treating dogwhistles as conventional implicatures.
- Most importantly, the dogwhistle effect is not conventionalized and is entirely listener-based.
- This preserves the speaker's deniability, a critical fact that CI accounts miss.

### Weaknesses of the Inferentialist View

- An account based entirely on extensional content and listener beliefs is too weak.
- As Khoo himself notes, the account wrongly predicts that any two coextensive terms should induce the same dogwhistles, but they don't.
- This requires a theory where inferences are tied to specific linguistic expressions, but are not part of their conventional semantic meaning.

### The Need for (some kind of) conventionalization

- The need to look at specific linguistic expressions undermines a purely content-based inferentialist theory.
- Recent work shows that dogwhistles pass tests for conventionalization in language.
- We must respect that dogwhistles are, in some sense, lexicalized / conventionalized.
- They are lexicalized, just not in the same way as CIs or standard truth-conditional content.
- The fact that dogwhistles are arbitrary, involving some amount of conventionaliztion, is enough to conclude a linguistic phenomenon is at hand, not just general reasoning.

## 2.3 The (Manner) Implicature View

### The (Manner) Implicature View

- In light of the previous views, we seem to have a paradox.
- We don't want conventionalized content (for deniability), but we must recognize that certain expressions are conventionally dogwhistles.
- A potential way out is to treat dogwhistles via Gricean implicature, specifically a Manner implicature.

### How Manner Implicatures Work

- Manner implicatures arise from using "non-standard" ways to say things.
- Listeners assume speakers will describe normal events with normal language.
- Utterance: "George made the car stop".
- This competes with the simpler, more normal "George stopped the car".
- The listener concludes something was non-standard about the action, generating the implicature that he stopped it in a weird way.

### A Manner Account of Dogwhistles?

- The idea: a dogwhistle like *inner-city* is not lexicalized, but is recognized as a non-standard way to say *Black neighborhoods*.
- This would trigger a Manner implicature, which would be deniable.
- The reasoning would be:

> The speaker said *inner-city*. Assuming she is cooperative...she would have said *Black neighborhoods*...

### Problem 1: The 'What is Said' Argument

- We must stop the reasoning immediately because of a problem.
- Classic Manner implicatures require that the two expressions mean the same thing (e.g., *stop the car* vs. *made the car stop*).
- But we have already seen that dogwhistles must not have this conventionalized content because they are deniable.
- This was the core argument against the CI view.

**Problem 2: The 'What is Implicated' Argument**

- Even if we grant the meaning equivalence, the implicature is still hard to derive.
- Normally, an asserted $S$ (meaning $P$) competes with $S'$ (also meaning $P$) to get an enriched meaning $P + E$.
- Here, we have *inner-city* ($S$) competing with *Black neighborhoods* ($S'$).
- The goal is to arrive at the enriched meaning *inner-city Black neighborhoods*, which essentially includes the alternative expression itself.
- This is radically different from the standard case; it would be like saying "made the car stop" to implicate that you "stopped the car," which is ridiculous.

## 2.4 The Speech Act View

**The Speech Act View**

- The Austinian motto: we must know not just what our utterances mean, but also what they *do*.
- Given the talk of subaudiences and covert moves, dogwhistles seem ideal candidates for a Speech Act Theory account.
- There is no such existing work on dogwhistles, but there is on adjacent phenomena like slurs and hate speech.

**Dogwhistles as Exercitives?**

- Research on slurs and hate speech treats them as having an *exercitive force*—they grant permission to the listener to hold certain views or engage in certain acts.
- Could some dogwhistles be covert exercitives?
- This seems plausible for racial dogwhistles that give a subaudience permission to indulge in racist attitudes.

**Problems for a Unified Speech Act View**

- It is hard to find a similar permission-granting effect in all dogwhistles.
- Take Bush's use of "wonder-working power" to signal to evangelical voters.
- The effect here is more about affecting solidarity or common spirit, not giving permission to act or speak in a certain way.
- Therefore, it is unlikely that dogwhistles in general can be identified with a single, previously known speech act like exercitives.

**Is 'Dogwhistling' a Speech Act?**

- A different possibility: perhaps dogwhistling is its *own* speech act, alongside commanding and promising.
- This view is much more amenable.
- Just as one needs a formal account of how an utterance gets the force of a promise, the analysis developed in this book explains how expressions come to act as dogwhistles.
- This work could be seen as providing a formal account of dogwhistle speech acts.

### 2.5 Mixed Views

**Mixed Views**

- More recent accounts involve novel or mixed types of meaning.
- Lo Guercio and Caso (2022) provide one such account.
- For them, dogwhistling is a two-step process.

**Lo Guercio y Caso's (2022) Two-Step Account**

1. The dogwhistle encodes a **conversational perspective**, which is treated as a not-at-issue conventional implicature (CI).
2. Upon receiving this perspective, the listener generates a **positioning message**, which can be an implicature, entailment, or perlocutionary inference.

- For example, "Dred Scott Decision" carries a CI that the speaker has an anti-abortion perspective.
- From this, a listener might infer the speaker will not appoint pro-choice judges.

**Critique of the Mixed View**

- Because this account relies on a CI for the first step, it is subject to the same objections as the pure CI view.
- Specifically, conventional content is not deniable, but dogwhistles are.
- Another problem: the same expression could trigger different perspectives depending on the speaker (e.g., pastor vs. politician).
- This would require assuming the expression is massively ambiguous across many dialects, which is not typical for NAI expressions.

## 3 The Social Meaning View

**Part 2: The Social Meaning View**

- Having surveyed previous accounts, we now present our own positive proposal.
- Our proposal: Dogwhistles involve particular kinds of inferences over the **social meaning** of expressions.

**What is Social Meaning?**

- Modern work on social meaning goes back to Labov's (1963) study of Martha's Vineyard English.
- In that study, the pronunciation of a diphthong ([ai]) became a marker in an ideological battle over the island's economy.
- The pronunciation became imbued with meaning, but not standard denotational meaning.
- At its core, the social meaning of an expression is the information it carries about how the speaker fits into the social landscape.
- It helps characterize an interlocutor's *persona*.

**A Social Meaning Theory of Dogwhistles**

A dogwhistle like Bush's "wonder-working power" has three key features:

1. It bears social meaning (i.e., it signals a persona).
2. Some members of the audience would disapprove of the speaker bearing that persona.
3. There is some uncertainty (or complete obliviousness) in the disapproving audience about whether the expression bears that social meaning.

**How this View Solves the Central Tension**

- This view resolves the paradox between the CI view and the inferentialist view.
- We don't want to associate dogwhistles with conventional truth-conditional content because they are deniable.
- But we must accept that they are conventional in some way, since synonymous expressions behave differently.
- **Solution**: Synonymous expressions can have the same truth-conditional content but differ in their *social meanings*. What is conventionalized is the social meaning, not a secondary proposition.

**Explaining Deniability**

- Social meanings are more mutable and context-dependent than truth-conditional meaning.
- For example, the pronunciation of "-ing" can signal a range of personas from uneducated to relaxed and unpretentious.
- It makes sense to treat social meaning as involving probabilistic associations between forms and meanings.
- With only a probabilistic link, there is room for a speaker to deny that the listener's inference about their social category was correct.
- This makes social meanings deniable in a way that is closer to conversational implicatures.

**Explaining Audience Asymmetry**

- The probabilistic turn also helps explain why some listeners fail to "hear" the dogwhistle.
- If social meanings are probabilistic, it is a simple step to allow that probability to vary across members of a speech community.
- Some speakers can be unaware that "wonder-working power" is commonly used by evangelicals, and thus is good evidence for that persona.
- This emerges from variation in what knowledge people have about how others use language.

**Can Social Meaning Affect Propositions?**

- Yes. While the approach suppresses the role of truth-conditional content, it can still trigger propositional inferences.
- We need a way for inferences about social meaning to affect the propositions a listener is willing to endorse.
- For example, hearing a Vineyarder use the local [ai] pronunciation signals they would likely agree to certain propositions about island life.
- The fact that dogwhistles have propositional effects does not preclude treating them as a social meaning phenomenon at their core.

# 4 Varieties of Dogwhistles

**Varieties of Dogwhistles**

- Having proposed that dogwhistles involve social meaning does not answer whether they are a unified phenomenon.
- In this section, we consider various typologies for dogwhistles that have been proposed.
- We will end by proposing our own typology.

## 4.1 Saul's 4-Way Typology

**Saul's (2018) 4-Way Typology**

Saul 2018 presents a typology based on two cross-classifying features:

1. **Intentional vs. Unintentional**: Whether the speaker *means* to dogwhistle.
2. **Overt vs. Covert**: How the dogwhistle is received by the audience.

**Intentional vs. Unintentional**

- This distinction is needed because dogwhistles can have an effect even when merely "mentioned," not "used".
- For example, a newscast replaying a dogwhistle-bearing ad to discuss the controversy can still generate the dogwhistle effect.
- This happens even if the newscaster did not intend to dogwhistle.

**Overt vs. Covert Dogwhistles**

This distinction concerns how the dogwhistle is received by the audience.

- **Overt Dogwhistles**: Correspond to the way we have mostly been discussing them. A savvy subaudience has declarative knowledge and can extract the extra information.
- **Covert Dogwhistles**: Seem to require a particular kind of ignorance in the target audience to have an effect.

**Example: Overt Dogwhistle**

- Albertson (2015) found that Pentecostals explicitly recognized the phrase "wonder-working power".
- Because of this declarative knowledge, they could identify a religious appeal in political speech that used the phrase.
- The savvy audience consciously extracts information that the unsavvy audience cannot.

**Example: Covert Dogwhistle**

- Mendelberg (2001) found that ads with racial dogwhistles shifted voting intentions of white voters high in racial resentment.
- **BUT**, if the dogwhistle was explicitly explained to these voters, the effect disappeared.
- For these dogwhistles, declarative knowledge *dissolves* the effect.
- Mendelberg explains this as implicit racial bias being triggered when the dogwhistle is not consciously processed, but this effect is overridden by consciously-held norms of racial equality when the dogwhistle is made explicit.

**Accommodating Saul's Typology**

- The overt/covert distinction can be handled by our social-meaning theory.
- The conventionalized social meaning of the expression doesn't change; what changes is how the listener reacts to that content.
- The distinction is at the level of listener uptake—whether it is conscious or unconscious, and how it interacts with other norms and beliefs.
- The four types emerge from how a uniform class of dogwhistles interacts with different kinds of participants in discourse.

## 4.2 Multivocalism

**Multivocalism**

- The political science literature often uses the term *multivocalism* instead of dogwhistling.
- This literature makes a distinction based on whether the speaker would deny the covert message.

- **Dogwhistles**: Covert messages the speaker *would deny*.
- **Multivocal Expressions**: Covert messages the speaker *would not deny*.

**Example: Multivocal Expression**

- George Bush's use of "wonder-working power" sent a risky message to evangelicals.
- However, if confronted, he would undoubtedly assent that he is an evangelical and believes in that power.
- Since he would not deny the covert message, this is classified as multivocalism.

**Example: Dogwhistle (Multivocalist view)**

- Paul Ryan's statement about "inner-cities" and the "culture of work".
- When called out, his office issued a denial, and he explicitly stated "This isn't a race based comment".
- Because Ryan denied the racial component of the message, this would be a bona fide dogwhistle under the multivocalist view.

**Why We Don't Adopt This Distinction**

- While an interesting distinction, we will not be making it.
- Our account can model both cases where a speaker would deny the message and cases where they would not.
- The reason is that there is no *linguistic* distinction to be made.
- There are no expressions that can *only* be dogwhistles or *only* be used in multivocal appeals.
- The distinction is about speaker strategy, not language, so we call them all dogwhistles.

## 4.3 Our Proposed Typology

**Our Typology: A Novel, Linguistic Distinction**

- We want to focus on distinctions that hew closely to the kinds of messages sent by dogwhistles.
- We draw a primary, novel distinction between two kinds of dogwhistles:
  1. **Identifying Dogwhistles**.
  2. **Enriching Dogwhistles**.

**Identifying Dogwhistles**

- **Function**: The content sends one message to all, while the dogwhistle transmits information about the speaker's true identity (persona) to a sub-audience.
- The covert content concerns the speaker's identity.

- **Example**: George Bush's use of "wonder-working power" doesn't change the meaning of "power".
- Instead, it flags him as an evangelical to those who recognize the phrase.

**Enriching Dogwhistles**

- **Function**: The content sends one message to all, while the whistle places an addendum on that message for a sub-audience.
- It involves a covert message, or *enrichment*, of what is said.
- **Example**: Paul Ryan's use of "inner city" conveys a geographical location to all audiences.
- But for a sub-audience, it specifically picks out and enriches the meaning to "Black neighborhoods" in those cities.

**A Cline, Not a Hard Boundary**

- These two types are likely extremes on a cline.
- Our analysis will show that **all** dogwhistles send information about the speaker's persona (i.e., they are all identifying).
- Whether the dogwhistle *also* exhibits enrichment depends on the personas, background assumptions, and at-issue content.
- Some uses lead to strong enrichments (like "inner-city"), while others lead to weak or absent enrichments (like "wonder-working power").

# 5 Summary & Conclusion

**Chapter Summary**

- We have presented the previous literature on dogwhistles and ruled out accounts using standard tools like conventional and Gricean implicatures.
- We advanced our own proposal: dogwhistles involve **social meaning** and the inferences that flow from it.
- We presented a new typology based on the message conveyed:
  - All dogwhistles convey information about the speaker's persona.
  - Some also provide enrichments to the truth-conditional content.

**Looking Ahead**

- Our unified, social meaning-based approach shows that even truth-conditional enrichments depend on retrieving social meaning.
- Treating dogwhistles as bearing social meaning means we must recapitulate pragmatic theory in this domain.

- To do so, we will move away from an explicitly Gricean framework, though there are proposals that RSA is in the Gricean spirit.
- The next chapter will introduce a pragmatic theory for social meaning in the Bayesian Rational Speech Act (RSA) framework.
- Subsequent chapters will build on this to create a comprehensive account of dogwhistles.

# Chapter 3: A Probabilistic Pragmatics for Dogwhistles

## 6 Introduction

**Chapter 2 Recap**

- The previous chapter situated dogwhistles within the wider field of pragmatics.
- We argued that dogwhistles involve pragmatic inferences over social conventions of language use.
- They are not expressions with conventionalized secondary meanings, but expressions used in particular ways by particular communities.
- This allows for in-group understanding and out-group deniability.

**Goal: A Unified Pragmatic Theory**

- While dogwhistles involve social meaning, a novel concept for formal semantics, we want the pragmatic mechanisms to be the same as those for truth-conditional content.
- This chapter aims to lay the foundation for a unified account of the pragmatics of both truth-conditional and social meanings.
- The result will be a satisfying analysis where dogwhistles, sociolinguistic inferences, and classical Gricean implicatures all belong to the same class of phenomena.

**Our Framework: Bayesian RSA**

- We will use the Bayesian Rational Speech Act (RSA) framework.
- RSA is a pragmatic theory that models the behavior of agents in signaling games.
- We will review the basics of RSA and then introduce Social Meaning Games (SMGs), a variant of RSA used for sociolinguistics.
- These formal ideas will recur throughout the rest of this work.

**The Advantage of a Game-Theoretic Approach**

- Bayesian RSA can sidestep the issue of conscious intention.
- In game theory, agent strategies can be interpreted in two ways:
  1. As conscious, rational decisions.
  2. As non-intentional convergence toward a utility-maximizing equilibrium (like in evolutionary biology).
- This flexibility makes it a suitable framework for phenomena that may operate below the level of consciousness.

**Interpreting Our Models**

- Our goal is not to build a comprehensive, explanatory model of a dogwhistling agent's mind.
- Rather, we use these game-theoretic models as an *epistemic tool*.
- The models help us tease out the core theoretical issues and pragmatic pressures that underlie dogwhistles.
- Instead of Gricean maxims, we specify speaker/hearer prior beliefs and utility functions.
- Pragmatic phenomena then emerge from rational, utility-maximizing behavior within the game's structure.

# 7 Bayesian Rational Speech Act Theory

**The Core of RSA**

- RSA captures the traditional divide between semantics (conventional meaning) and pragmatics (meaning in context).
- It models how semantic meaning is enriched through a process of speakers and listeners recursively reasoning about each other's conversational moves.

**The RSA Hierarchy of Reasoning**

The framework is built on levels of increasing pragmatic sophistication:

- **The Literal Listener ($L_0$)**: Does no pragmatic reasoning. Interprets utterances based only on their conventional, semantic meaning. This is the "ground floor" of semantics.
- **The Pragmatic Speaker ($S_1$)**: Reasons about $L_0$. Chooses an utterance to best help a literal listener identify the true state of the world.
- **The Pragmatic Listener ($L_1$)**: Reasons about $S_1$. Interprets an utterance by thinking about why a pragmatic speaker ($S_1$) would have chosen it.

This process can continue ($S_2, L_2$, etc.), but the first three layers can already model many interesting phenomena.

## 7.1 The Literal Listener L0

**The Literal Listener ($L_0$)**

- The literal listener, $L_0$, hears an utterance $u$ and updates her beliefs about the probability of each possible world $w$.
- She does this using vanilla Bayesian inference.
- An utterance's conventional meaning, $[\![u]\!]$, is a proposition (a function from worlds to truth values).

The probability $L_0$ assigns to a world $w$ given utterance $u$ is:

$$P_{L_0}(w|u) \propto [\![u]\!](w) \times P(w)$$

This is proportional to the utterance's truth value in that world (1 or 0) multiplied by the prior probability of that world.

### $L_0$ Example: Setup

- Imagine a world with three possibilities:
  - $w_p$: a world where only proposition $p$ is true.
  - $w_q$: a world where only proposition $q$ is true.
  - $w_\emptyset$: a world where both are false.
- Assume $L_0$ starts with a uniform prior belief over these worlds:

$$P(w_p) = P(w_q) = P(w_\emptyset) = \frac{1}{3}$$

### $L_0$ Example: Receiving an Utterance

Now, $L_0$ hears the utterance $u_{p \vee q}$ ("p or q"). She computes the posterior probability for each world:

- For $w_p$: $P_{L_0}(w_p|u_{p\vee q}) \propto [\![p \vee q]\!](w_p) \times P(w_p) = 1 \times \frac{1}{3}$
- For $w_q$: $P_{L_0}(w_q|u_{p\vee q}) \propto [\![p \vee q]\!](w_q) \times P(w_q) = 1 \times \frac{1}{3}$
- For $w_\emptyset$: $P_{L_0}(w_\emptyset|u_{p\vee q}) \propto [\![p \vee q]\!](w_\emptyset) \times P(w_\emptyset) = 0 \times \frac{1}{3}$

The utterance is informative: it rules out $w_\emptyset$ entirely.

### $L_0$ Example: Final Beliefs

- The probability mass from the eliminated world ($w_\emptyset$) is redistributed over the remaining worlds.
- Because the initial weights for $w_p$ and $w_q$ were equal, they now share the total probability.
- $L_0$'s new (posterior) beliefs are:

$$P(w_p) = P(w_q) = \frac{1}{2}$$

$$P(w_\emptyset) = 0$$

- This shows how a literal listener can probabilistically update beliefs based on conventional meaning.

## 7.2 The Pragmatic Speaker S1

**The Pragmatic Speaker ($S_1$)**

- A cooperative speaker is an epistemic ally; she chooses utterances to help a listener know the state of the world.
- In RSA, this means the speaker ($S_1$) chooses an utterance $u$ to maximize the probability that a listener ($L_0$) will correctly infer the world $w$.
- In other words, the speaker's utility is a function of how the listener interprets her messages.

**Speaker Utility**

The utility for speaker $S_1$ of choosing utterance $u$ in world $w$ is:

$$U_{S_1}(u, w) \propto \ln(P_{L_0}(w|u)) - C(u)$$

- The first term, $\ln(P_{L_0}(w|u))$, captures how informative the utterance is for the literal listener.
- The second term, $C(u)$, is the **cost** of the utterance.

**The Cost Function $C(u)$**

- Pragmatic reasoning doesn't happen against all possible alternative utterances. The space of alternatives is constrained.
- The cost function $C(u)$ implements this idea.
- It can represent various factors:
  - **Complexity**: "Bill ate three cookies and no more" is more complex (higher cost) than "Bill ate three cookies".
  - **Politeness**: A rude utterance has a higher social cost.
  - **Novelty**: An unusual phrasing might be more costly.
- In the context of dogwhistles, this can model the social cost of listener disapproval.

**Speaker Choice: The Softmax Function**

- We don't assume speakers are perfectly rational. They might sometimes choose a sub-optimal utterance.
- We model this by saying speakers choose utterances probabilistically, based on utility. This is often done with a "softmax" function.

The probability that speaker $S_1$ chooses utterance $u$ in world $w$ is:

$$P_{S_1}(u|w) \propto \exp(\alpha \times U_{S_1}(u, w))$$

The parameter $\alpha$ is a "rationality" or "temperature" parameter (actually I think officially an inverse-temperature parameter).

**The Rationality Parameter $\alpha$**

$$P_{S_1}(u|w) \propto \exp(\alpha \times U_{S_1}(u, w))$$

The $\alpha$ parameter acts as a control knob for speaker rationality:

- As $\alpha \to \infty$, the speaker becomes perfectly rational, always choosing the utterance with the highest utility.
- As $\alpha \to 0$, the speaker becomes more random, choosing more non-utility-maximizing utterance.
- Intermediate values of $\alpha$ allow us to model a "boundedly rational" speaker who is more likely to choose better options, but doesn't always choose the single best one.

## 7.3 The Pragmatic Listener L1

**The Pragmatic Listener ($L_1$)**

- The pragmatic listener $L_1$ does the same thing as the literal listener $L_0$: uses Bayesian inference to determine the most likely world.
- The crucial difference is what the inference is based on.
- Instead of using the conventional semantics directly, $L_1$ reasons about the actions of a pragmatic speaker, $S_1$.

The probability $L_1$ assigns to world $w$ given utterance $u$ is:

$$P_{L_1}(w|u) \propto P_{S_1}(u|w) \times P(w)$$

**Unraveling the Recursion**

The elegance of the RSA framework masks its beautiful recursive complexity. When a pragmatic listener $L_1$ hears an utterance, she is reasoning:

1. "What is the most likely state of the world, given that..."
2. "...a (boundedly rational) pragmatic speaker $S_1$ chose this utterance for me..."
3. "...because she was modeling how a literal listener $L_0$ would interpret it based on its conventional meaning..."
4. "...and she was weighing the utterance's informativity against its cost."

**The Full $L_1$ Formula**

Unpacking the recursive definitions gives us the full formula for the pragmatic listener's inference:

$$P_{L_1}(w|u) \propto \exp(\alpha \times (\ln([\![u]\!](w) \times P(w)) - C(u))) \times P(w)$$

This single formula encapsulates the entire RSA framework from the listener's perspective, showing how conventional meaning ($[\![u]\!](w)$) is pragmatically enriched through reasoning about a boundedly rational, cost-sensitive speaker.

# 8 Case Study: Implicature in RSA

### Case Study: Scalar Implicature

- Let's see how RSA models the classic scalar implicature "some but not all".
- **Scenario**: There are three croissants. You ask if I've eaten any.
- **My Reply**: "I've eaten some."
- **Your Inference**: You understand that I have not eaten all of them, even though I didn't explicitly say so.

### The Classic Gricean Story

Your reasoning process is explained as follows:

1. The speaker said "some".
2. Why didn't they say "all"? "All" is relevant, stronger, and not more complex.
3. The only reason must be that they do not believe "all" is true.
4. Since the speaker knows how many they ate, they must believe "all" is false.
5. Therefore, "some" implicates "not all".

### The RSA Model Setup

- **Worlds**: 4 possible states of the world.
  - $w_0$: ate 0 croissants
  - $w_1$: ate 1 croissant
  - $w_2$: ate 2 croissants
  - $w_3$: ate 3 croissants (i.e., "all")
- **Utterances**: 3 possible messages.
  - NONE (true in $w_0$)
  - SOME (true in $w_1, w_2, w_3$)
  - ALL (true in $w_3$)
- **Priors**: We assume a uniform prior over worlds and equal costs for utterances.

### The Literal Listener $L_0$

The literal listener interprets "some" as being true in any world where at least one croissant was eaten.

**$L_0$ Posterior Probabilities**

| Utterance | $P(w_0\|u)$ | $P(w_1\|u)$ | $P(w_2\|u)$ | $P(w_3\|u)$ |
|---|---|---|---|---|
| SOME | 0 | 0.33 | 0.33 | 0.33 |
| ALL | 0 | 0 | 0 | 1 |

To $L_0$, hearing "some" makes worlds $w_1, w_2$, and $w_3$ equally likely.

### The Pragmatic Speaker $S_1$

The pragmatic speaker $S_1$ chooses the most informative utterance for $L_0$.

- In $w_1$ or $w_2$, she must say SOME.
- **The critical state is $w_3$ (ate all)**: Here, both SOME and ALL are true.
- ALL is more informative for $L_0$ (it identifies $w_3$ perfectly), so a rational speaker will strongly prefer it.

**$S_1$ Choice Probabilities**

| World | $P(u = \text{NONE})$ | $P(u = \text{SOME})$ | $P(u = \text{ALL})$ |
|---|---|---|---|
| $w_0$ | 1 | 0 | 0 |
| $w_1$ | 0 | 1 | 0 |
| $w_2$ | 0 | 1 | 0 |
| $w_3$ | 0 | 0.25 | 0.75 |

### The Pragmatic Listener $L_1$

Now we get to the crux: how does the pragmatic listener $L_1$ interpret the utterance SOME?

- $L_1$ reasons about why $S_1$ chose SOME.
- $L_1$ sees from the $S_1$ table that if the world were actually $w_3$ (all), it is relatively *unlikely* that $S_1$ would have said SOME. She would have preferred the more informative ALL.
- This fact guides $L_1$ to assign a lower probability to $w_3$.

### The Implicature Emerges

When $L_1$ hears SOME, she updates her beliefs based on the $S_1$ probabilities:

- The probability of $w_3$ (all) is low: $P_{L_1}(w_3|\text{SOME}) \approx .11$
- The probabilities of $w_1$ or $w_2$ (some but not all) are high: $P_{L_1}(w_1|\text{SOME}) = P_{L_1}(w_2|\text{SOME}) \approx .44$

The listener is now much more confident that the world is $w_1$ or $w_2$, not $w_3$. **This is the scalar implicature**: "some" has been pragmatically enriched to mean "some but not all".

## 9 Social Meaning Games

### Extending RSA: Social Meaning Games

- The RSA framework is modular. It can be extended to reason over other kinds of meaning besides truth conditions.
- In influential work, Burnett (2017, 2019) does precisely this, creating a formal foundation for variationist sociolinguistics.
- This extension allows us to model how speakers use language to construct social identity.

### 'Third Wave' Sociolinguistics

The core idea of 'Third Wave' sociolinguistics is that linguistic variation is not just a product of pre-existing, static social identities.

- People don't just get assigned an identity and then speak accordingly.
- Instead, sociolinguistic practice is deeply creative.
- Through their linguistic choices, speakers constantly create a place for themselves in social space.
- Variation is the raw material from which speakers construct, entrench, and mutate social identities through their stylistic practices.

### The Indexical Field

- Eckert (2008) calls this raw material the **indexical field**.
- Consider the contrast between the apical [in] and velar [iŋ] pronunciations of "-ing".
- These variants are associated with a range of contrasting properties:

uneducated ∼ educated
relaxed ∼ formal
lazy ∼ effortful
unpretentious ∼ pretentious

- These oppositions are not simple binaries; they form a complex field of potential social meanings.

### From Indexical Fields to Personas

- Which properties a hearer assigns to a speaker depends on the variant, the speaker's entire style, and the context.
- Speakers select from variants to construct a style that conveys a particular **persona**.
- These personas (e.g., "diva," "jock," "nerd") are constructed in part through style and are linked to our common ideas of sociocultural identities.

### Example: Heath, the 'Diva'

- Podesva (2004, 2007) discusses Heath, a gay medical student who modulates his phonetic production throughout the day.
- At a BBQ with friends, he exaggerates his stop-releases and uses more falsetto to construct a "diva" persona.
- How does this work? The "diva" stereotype is associated with being demanding, emotive, and fussy.
- The acoustic features Heath uses (exaggeratedly releasing stops, using falsetto) have social meanings like "effortful," "pretentious," and "emotive" in their indexical fields.
- The style is coherent with the persona he wants to construct.

### Adding Indexical Fields to RSA

Burnett (2017, 2019) translates these ideas into the RSA framework.

- An expression $u$ has not just a truth-conditional meaning $[\![u]\!]$, but also a **social meaning** $[u]$.
- The social meaning $[u]$ is its indexical field: a set of properties from a universe of social properties $\mathbb{P}$.
- A **persona** $\pi$ is defined as a maximally consistent set of these properties.

### The Eckert-Montague Field

- The connection between an utterance's social meaning and the personas it can signal is called its **Eckert-Montague field (emf)**.
- The emf of an utterance $u$ is the set of all personas that are consistent with its social meaning $[u]$.

$$[\mathbf{emf}(u)](\pi) = 1 \text{ iff } \pi \cap [u] \neq \emptyset$$

- When a speaker chooses a variant, she projects a set of possible personas she could be assigned, ruling out those inconsistent with the variant's social meaning. This is the setup for the game.

### Sociolinguistic Speakers and Listeners

The game is parallel to the standard RSA game, with a few key differences:

- **What is being signaled?**: Not states of the world, but the speaker's intended **persona**.
- **Literal Listener ($L_0$)**: Reasons about which personas are consistent with an utterance's social meaning (its emf).
- **Speaker ($S_1$)**: Chooses variants to help the listener identify her desired persona.

### A Key Difference: Speaker Utility

- In standard pragmatics, the speaker observes the world and reports on it.
- In sociolinguistics, the speaker doesn't just report their identity; they actively *construct* it.
- Therefore, the speaker's utility must also include her own preference for projecting a certain persona.

We add an **affective value** function, $\nu(\pi)$, representing the speaker's preference for persona $\pi$.

$$U_{S_1}(u, \pi) \propto \ln(P_{L_0}(\pi|u)) - C(u) + \nu(\pi)$$

# 10 Case Study: Burnett on Obama

### Case Study: Barack Obama's '-ing'

- Burnett (2019) applies this model to Barack Obama's use of apical [in] vs. velar [iŋ].
- **Observation** (Labov, 2012):
  - At a casual Father's Day BBQ: used apical [in] **72%** of the time.
  - Taking questions with reporters after the BBQ: used apical [in] only **33%** of the time.
- This shows he is controlling his use of variants based on the social situation and the persona he wants to project.

### The Model Setup: Personas

Burnett uses a simplified model:

- **Indexical Fields**:

  iŋ : {competent, aloof}

  in : {incompetent, friendly}

- **Four Possible Personas**:
  - **Stern Leader**: {competent, aloof}
  - **Cool Guy**: {competent, friendly}
  - **Doofus**: {incompetent, friendly}
  - **Asshole**: {incompetent, aloof}

Obama's task is to choose variants to come off as the "Cool Guy" at the BBQ and the "Stern Leader" at the press conference.

### Speaker Preferences (The BBQ Context)

At the BBQ, Obama wants to be seen as a "Cool Guy." We model this with the affective value function $\nu$.

**Obama's Preferences ($\nu$)**

| Persona | Value |
|---|---|
| Cool Guy | 1 |
| Stern Leader | 0.5 |
| Doofus | 0 |
| Asshole | 0 |

He strongly prefers the "Cool Guy" persona, is okay with "Stern Leader," and wants to avoid the other two.

### Listener Priors

- The speaker must also model the listener's prior beliefs.
- Let's say Obama is worried about a listener who thinks he is naturally aloof and might interpret his "Stern Leader" persona as just being an "Asshole" at a party.

**A Listener Biased Towards 'Aloof'**

| Persona | Prior Probability |
|---|---|
| Cool Guy | 0.2 |
| Stern Leader | 0.3 |
| Doofus | 0.2 |
| Asshole | 0.3 |

### Listener's Interpretation ($L_0$)

Given these priors, a literal listener will update their beliefs as follows upon hearing a variant:

**Listener Posteriors ($L_0$)**

| Variant | Cool Guy | Stern Leader | Doofus | Asshole |
|---|---|---|---|---|
| -iŋ | 0.25 | 0.375 | 0 | 0.375 |
| -in | 0.286 | 0 | 0.286 | 0.428 |

- Using the velar -iŋ confirms the listener's belief that Obama is aloof.
- Using the apical -in rules out the "Stern Leader" persona and provides more evidence that he is not trying to be aloof.

### Obama's Utility Calculation

By combining the listener's posteriors with Obama's own preferences ($\nu$), we can calculate the utility for each choice-persona pair.

**Obama's Utility ($\mathbf{U_{S_1}}$)**

| Target Persona | Variant | Utility |
|---|---|---|
| Cool Guy | -in | $-0.253$ |
| Cool Guy | -iŋ | $-0.386$ |
| Stern Leader | -iŋ | $-0.481$ |
| ... (other pairs less optimal) | | ... |

The top line result shows that the best strategy to achieve the "Cool Guy" persona is to use the apical -in variant.

### Matching the Data

- The model shows that a perfectly rational Obama should use the apical -in variant to try to construct the "Cool Guy" persona.
- By setting the rationality parameter $\alpha$ to 8 (modeling a boundedly rational speaker), the model predicts the speaker will choose the apical variant 72

- This perfectly matches the observed data from Labov's study of Obama at the BBQ.
- This demonstrates the power of Social Meaning Games to model real-world sociolinguistic interactions.

# 11 Conclusion

**The Power of Social Meaning Games**

- Burnett's Social Meaning Games provide a powerful framework for modeling sociolinguistic behavior.
- A major advance is the promise of unifying formal pragmatics (reasoning about truth) with Third Wave variationist sociolinguistics (reasoning about style and identity).
- It also provides a formal model of social agents that can be used to generate testable predictions about human behavior in real-world scenarios.

**Next Steps: Back to Dogwhistles**

- While we wholeheartedly endorse this approach, we cannot use it to model dogwhistles without further modification.
- The next chapter will make the minimal modifications necessary to begin to understand how dogwhistles work.
- This provides a solid starting point with clear connections back to the established literature on pragmatics and sociolinguistics.

# Chapter 4: Identifying Dogwhistles

## 12 Introduction

### Chapter 3 Recap: Social Meaning Games

- The previous chapter introduced the game-theoretic account of sociolinguistic interaction from Burnett 2017, 2019.
- When variants exist, social meaning can accrete to them.
- Speakers choose variants to construct a style that signals their desired persona.
- Listeners reason about the social meanings of these variants to situate the speaker in social space.

### Connecting to Dogwhistles

Thinking back to Albertson's (2015) work on religious dogwhistles, the connections become clear.

> **Example: Bush's "Wonder-Working Power"**
> "Yet there's power—wonder-working power—in the goodness and idealism and faith of the American people."

- Religious voters were more likely to recognize the phrase and support the candidate.
- Those who recognized it but opposed religious appeals were less likely to approve.
- This process is mediated by partisan identity; listeners who detect such dogwhistles are more likely to label the speaker as Republican.

### Identifying Dogwhistles

- This interaction looks like a more complicated version of standard sociolinguistic signaling.
- The speaker (Bush) wants to convey a conservative Republican persona, but only to listeners who would approve.
- He picks variants whose social meaning will guide a subset of listeners to assign him that persona, while leading others away from it.
- This is our proposal for **identifying dogwhistles**: standard sociolinguistic inference, but performed strategically in an adversarial setting.

### Chapter Goals

1. Understand the contextual conditions that make identifying dogwhistles possible.
2. Understand when it is rational for speakers to use them.
3. Understand what kind of language makes for a good identifying dogwhistle.

To achieve this, we must extend the formal model of sociolinguistic variation from the previous chapter.

# 13 A Model for Dogwhistles

### Extending Social Meaning Games

- At the heart of Burnett's Social Meaning Games (SMGs) is a literal listener $L_0$ who is completely aware of the sociolinguistic conventions of the language.
- Her inference is based on which personas an utterance is consistent with:

$$L_0(p|m) \propto [\mathbf{emf}(m)](p) \times P(p)$$

- We must step away from this assumption to model dogwhistles.

### The Core Modification: Listener Uncertainty

- The whole point of the "wonder-working power" dogwhistle is that some listeners recognize its connection to an evangelical persona, while others do not.
- We don't want to treat social meanings as fixed, lexical facts known by all.
- We must allow for **uncertainty** about which expressions flag which personas.
- This means listeners have priors not just over personas ($P(p)$), but also over the link between messages and personas ($P(m|p)$).

### A New Literal Listener

We redefine the literal listener to incorporate this uncertainty. The listener's inference is now proportional to the likelihood of the message given the persona:

$$L_0(p|m) \propto P(m|p) \times P(p)$$

- This is a generalization of Burnett's model.
- Dogwhistles can emerge when one group of listeners has a high value for $P(m|p)$ (a tight connection) while another group has a low value (unaware).

### The Adversarial Context: Listener Preferences

- Dogwhistles emerge in adversarial contexts.
- Speakers exploit variation in listener knowledge to get one group (approvers) to assign a persona while getting another group (disapprovers) to assign a more innocuous one.
- This means we must explicitly model listener preferences.

### Modeling Preferences: Affective Value

- We assume the utility calculation takes into account the message's social value.
- The speaker has a preference function $\nu_S$ that assigns a value to each persona.
- We introduce a second function for listeners, $\nu_L$, that represents their approval or disapproval of each persona.
- Speakers want to present themselves in a certain way, but are also sensitive to whether listeners will approve of that persona.

### A New Speaker Utility Function

The utility of a message $m$ is calculated based on the range of personas consistent with it and how a listener $L_0$ is likely to receive them.

$$U_{S_1}^{Soc}(m, L_0) = \sum_{p \in [m]} (P_{L_0}(p|m))$$

$$+\nu_{S_1}(p) P_{L_0}(p|m) + \nu_{L_0}(p) P_{L_0}(p|m)$$

This utility function combines:

- The informativity of the message.
- The speaker's preference for the resulting persona ($\nu_S$).
- The listener's approval of the resulting persona ($\nu_L$).

### Utility in a Group Setting

- Dogwhistles come into their own when speakers address groups with mixed preferences and beliefs.
- The simplest way to model utility for a group $G$ is to sum the utilities for each individual listener in the group:
$$U_{S_n}^{Soc}(m, G) = \sum_{L_{n-1} \in G} U_{S_n}^{Soc}(m, L_{n-1})$$
- This is a simplification; a speaker might care more about the reaction of a powerful individual, but it's a useful starting point.

## 14 Case Study: Jill Stein

### Case Study: A Predicament for Jill Stein

**The Scenario (circa 2018)**

Jill Stein is asked about vaccines. She knows her base is anti-corporate, but contains a passionate anti-vaxxer minority that others in her party dislike. The question itself implies uncertainty about her stance. She realizes this is a perfect occasion for a dogwhistle.

### The Messages and Personas

- Stein is choosing between messages that mark her as anti-corporate, but either pro- or anti-vaccine.
- **The Dogwhistle**: "big pharma". This is common in anti-vaxxer discourse.
- **Salient Alternative**: "corporate scientists". This is also anti-corporate but has no traction in the anti-vaxxer community.

**Listener Types and Priors**

- **Audience**: Passionate anti-vaxxers, clueless pro-vaxxers, and knowledgeable (savvy) pro-vaxxers.
- **Priors**: We assume all listeners have the same priors about Stein. They believe she is probably anti-corporate, but are unsure if she is pro- or anti-vaxxer.

**Priors for Stein's persona**

| | |
|---|---|
| {pro-vax, pro-corporate} | .05 |
| {pro-vax, anti-corporate} | .40 |
| {anti-vax, pro-corporate} | .15 |
| {anti-vax, anti-corporate} | .40 |

**Listener Preferences ($\nu_L$): Savvy Listeners**

The anti-vaxxers and savvy pro-vaxxers are highly motivated and polarized. They are mirror images of each other.

**Anti-vaxxers**

| Persona | Value |
|---|---|
| pro-vax, pro-corp | -125 |
| pro-vax, anti-corp | -100 |
| anti-vax, pro-corp | 75 |
| anti-vax, anti-corp | 100 |

**Savvy Pro-vaxxers**

| Persona | Value |
|---|---|
| pro-vax, pro-corp | 75 |
| pro-vax, anti-corp | 100 |
| anti-vax, pro-corp | -125 |
| anti-vax, anti-corp | -100 |

**Listener Preferences ($\nu_L$): Unsavvy Pro-Vaxxer**

The unsavvy pro-vaxxer is less invested. They are happy if Stein is pro-vax (a default position) but highly object if she is anti-vax.

**Unsavvy Pro-vaxxers**

| Persona | Value |
|---|---|
| pro-vax, pro-corp | 50 |
| pro-vax, anti-corp | 75 |
| anti-vax, pro-corp | -125 |
| anti-vax, anti-corp | -100 |

**Speaker Preferences ($\nu_S$): The Amoral Politician**

- For this example, we assume Stein is a completely amoral politician.
- She has no personal preference for any persona ($\nu_S(p) = 0$ for all $p$).

- Her only goal is to maximize her audience's reception.
- This framework allows us to model such extreme strategies, which are common in political discourse where speakers try to be "all things to all people."

**Listener Awareness ($P(m|p)$): The Key to the Dogwhistle**

This is the crucial parameter. Listeners differ in how they connect messages to personas.

- **Savvy Listeners** (both pro- and anti-vax) know "big pharma" is tightly linked to the anti-vaxxer/anti-corporate persona.

**Likelihood of "big pharma" for savvy listeners (both Pro- and Anti-vax)**

| personas | $Pr(m\|p)$ |
|---|---|
| {pro-vaxxer, pro-corporate} | 0 |
| {pro-vaxxer, anti-corporate} | .1 |
| {anti-vaxxer, pro-corporate} | .1 |
| {anti-vaxxer, anti-corporate} | .8 |

- **Unsavvy Listeners** don't see this tight connection; they interpret "big pharma" as primarily just an anti-corporate phrase.

**Likelihood of "big pharma" for unsavvy listeners**

| personas | $Pr(m\|p)$ |
|---|---|
| {pro-vaxxer, pro-corporate} | 0 |
| {pro-vaxxer, anti-corporate} | .4 |
| {anti-vaxxer, pro-corporate} | .1 |
| {anti-vaxxer, anti-corporate} | .2 |

- **"Corporate scientists" for savvy Listeners** interpret "corporate scientists" in the same way: as primarily pro-vax and anti-corporate.
- In particular, we take this phrase to all but rule out being anti-vaxxer, but to lean anti-corporate. The reason is that anti-vaxxers already have in-group language to disparage the sector (which savvy pro-vaxxers also know), yet the phrase does not convey positive affect.

**Likelihood of "corporate scientists" for savvy listeners**

| personas | $Pr(m\|p)$ |
|---|---|
| {pro-vaxxer, pro-corporate} | .6 |
| {pro-vaxxer, anti-corporate} | .8 |
| {anti-vaxxer, pro-corporate} | .1 |
| {anti-vaxxer, anti-corporate} | .1 |

- **Unsavvy likelihoods for "corporate scientists"**—For unsavvy listeners "corporate scientists" seems fairly bland. It is not more anti-corporate than "big pharma", but all things being equal, it is probably the way anti-corporate people speak. It it is not particularly linked to anti-vaxx rhetoric, which these listeners do not know much about.

**Likelihood of "corporate scientists" for unsavvy listeners**

| personas | $Pr(m\|p)$ |
|---|---|
| {pro-vaxxer, pro-corporate} | .3 |
| {pro-vaxxer, anti-corporate} | .5 |
| {anti-vaxxer, pro-corporate} | .1 |
| {anti-vaxxer, anti-corporate} | .2 |

### Calculated Utilities: Savvy Pro-Vaxxer

For a savvy pro-vaxxer, using "big pharma" is a bad move. They detect the dogwhistle and disapprove of the anti-vaxxer persona it signals.

**Utility for Stein (Savvy Pro-Vaxxer)**

| Message | Utility |
|---|---|
| big pharma | -85 |
| corporate scientists | 62 |

Stein is much safer using "corporate scientists" with this listener type.

### Calculated Utilities: Unsavvy Pro-Vaxxer

The unsavvy pro-vaxxer has a much higher utility for "big pharma." This is the dogwhistle effect in action.

**Utility for Stein (Unsavvy Pro-Vaxxer)**

| Message | Utility |
|---|---|
| big pharma | 32 |
| corporate scientists | 40 |

Because this listener doesn't hear the whistle, the negative impact is greatly reduced.

### Calculated Utilities: Savvy Anti-Vaxxer

The anti-vaxxer shows the opposite pattern from the savvy pro-vaxxer. They also hear the dogwhistle, but they strongly endorse its message.

**Utility for Stein (Savvy Anti-Vaxxer)**

| Message | Utility |
|---|---|
| big pharma | 73 |
| corporate scientists | -84 |

### When is it Optimal to Dogwhistle?

- The model already shows the identifying dogwhistle effect: a message's utility can be greatly increased when listeners fail to realize how tightly it's correlated with a persona they disapprove of.
- The model also makes predictions about when it is rational to deploy a dogwhistle based on audience composition.
- In a mixed audience, the ratio of anti-vaxxers to pro-vaxxers (especially savvy ones) determines the best strategy.

$(x_1 \times U_S^{Soc}(\text{DOGWHISTLE}, L^1)), ..., (x_n \times U_S^{Soc}(\text{DOGWHISTLE}, L^n)) >$
$(x_1 \times U_S^{Soc}(\text{DISAVOWAL}, L^1)), ..., (x_n \times U_S^{Soc}(\text{DISAVOWAL}, L^n))$

**Audience Composition Matters**

- If Stein is talking to any number of pro-vaxxers, she is best to issue a disavowal (use "corporate scientists").
- If she is talking only to anti-vaxxers, she should issue the dogwhistle.
- **Example**: In an audience of 5000, if she believes there are 400 anti-vaxxers but only 100 savvy pro-vaxxers, it would be optimal to use the dogwhistle.
- If the ratios change (e.g., adding 200 more savvy pro-vaxxers), the dogwhistle may no longer be a safe or rational choice.

# 15 Exploring the Parameters

**How Brittle is the Account?**

- Does this result depend on fine-tuning all the numerical parameters?
- The answer is no. The utility of dogwhistling emerges under a fairly broad and plausible space of parameter settings.
- Moreover, varying the parameters allows us to model different, interesting kinds of social agents.

**Parameter: Speaker's Preference ($\nu_S$)**

- The case study assumed a malleable speaker with no preferences, the type most likely to use dogwhistles to be "all things to all people."
- If a speaker has a preference for the dogwhistled persona, they will be willing to use it in more socially dangerous situations (i.e., with fewer allies in the audience).
- If a speaker dislikes the persona, they can be peer-pressured into using it only when many supportive in-group members are present.
- A speaker who absolutely abhors the persona (e.g., an anti-racist activist and a racist dogwhistle) will never use it.

**Parameter: Listener's Preference ($\nu_L$)**

- What matters here is the structure of the population as a whole.
- Dogwhistling is optimal as long as there are two groups of listeners with opposing preferences for a persona, and one group is less aware of the message that signals it.
- This setup is not contrived; it is likely the norm for in-group language. In-group members like each other and know their own rhetoric, while disapproving out-group members often don't associate with them and thus don't know their rhetoric.

**The 'Policing' Listener**

- A special listener type can mitigate dogwhistling.
- This is a listener who, like the in-group, is aware of the link between a message and a persona, but who assigns that persona a large negative value.
- These are listeners who actively oppose the in-group, rather than just disapproving of it. They monitor the group's rhetoric.
- Their presence can overwhelm the utility gained from the in-group, making the dogwhistle too costly, unless they are sufficiently outnumbered by other listener types.

**Parameter: Listener Priors ($P(p)$)**

- Dogwhistling requires audience uncertainty about the speaker's persona.
- Strong priors blunt the effect of new evidence (the message).
- If the audience has a strong prior that a speaker *cannot* bear a certain persona, then a dogwhistle for that persona will fail.
- **Example**: An NGO named "Inner-City Youth Arts." Their work gives them a very low prior for being racist agents. They can use the phrase "inner-city" without sending a covert racist signal.

**The Effect of Strong In-Group Priors**

- What if the in-group already has a strong prior that the speaker is "one of them"?
- In this case, dogwhistles can actually lose their utility. The in-group signaling effect is already baked in.
- The speaker is then freer to use more neutral terms to avoid alienating the out-group, knowing the in-group will discount the neutral language.
- **Example**: White supremacists interpreting Trump's disavowal of racism as something he "just has to say for cover." Their strong prior that he is an ally allows them to ignore the contrary evidence of the disavowal.

# 16 The Grammar of Dogwhistles

**What Makes a Good Identifying Dogwhistle?**

- We now ask: what kind of language is likely to have these properties?
- What are the linguistic characteristics of a good identifying dogwhistle?

**What Makes a BAD Dogwhistle: In-Group Slang**

- At first glance, dogwhistles might seem like mere in-group slang.
- But canonical slang makes for a poor dogwhistle because it fails the deniability test in adversarial contexts.
- If an out-group listener hears a term they don't understand at all, they can't assign it to any persona. This leads to a state of conversational absurdity.
- The listener knows a secret message is being sent, and the dogwhistle attempt is no longer covert.
- Experimental work confirms that listeners punish brands when they detect they are not the target audience.

**What Makes a GOOD Dogwhistle**

- The model predicts that novel expressions and pure in-group coinages make poor identifying dogwhistles.
- This shapes the kinds of dogwhistles we actually see.
- Good dogwhistles are often low-frequency expressions that are not completely novel and provide cover.

**Good Dogwhistle Source 1: Quotations**

- Quotations are ideal, especially snippets of larger works.
- They can be composed of entirely ordinary language, yet gain significance for a population that recognizes the source.
- **Example**: "wonder-working power." The words themselves are socially inert. For an average listener, they are consistent with any persona.
- But it is an instantly recognizable quote to evangelical Christians from the well-known hymn "There is Power in the Blood."
- This creates a powerful identifying dogwhistle.

**Good Dogwhistle Source 2: Memes and Named Entities**

- The accretion of social meaning through repetition is the core of memetics.
- **Example**: Gritty, the mascot for the Philadelphia Flyers. In the wider culture, he is just a mascot. On the far-left internet, he became a symbol of Antifa.
- A speaker can then use Gritty to express a centrist opinion (e.g., "Gritty says pay the players") while simultaneously constructing a far-left persona for themselves, an inference unavailable to the average listener.

**Example: Named Entity Dogwhistles**

- The Subaru ad campaign admitted to using dogwhistles to target lesbians.
- The cars in the ads had license plates like P TOWN and XENA LVR.
- P TOWN refers to Provincetown, MA, a well-known LGBTQ vacation spot.
- XENA LVR refers to "Xena: Warrior Princess," a 90s TV show that was a major cultural touchstone in the lesbian community.
- These references to people and places associated with the community serve as clear identifying dogwhistles.

# 17 Non-Linguistic Dogwhistles

**A Final Argument for Social Meaning**

- Why treat dogwhistles as trafficking in social meaning instead of some kind of secondary propositional content?
- One strong argument is the existence of completely non-linguistic identifying dogwhistles.

**Example: The Fred Perry Polo Shirt**

- The black Fred Perry polo with yellow stripes has a long history with UK working-class subcultures, some of which have overlapped with racist/fascist politics.
- This complex history makes it a perfect dogwhistle. One can wear it to signal they are a skinhead, but if called out, deny it by pointing to its connection to punk music.
- In 2020, it became the de-facto uniform of the Proud Boys.
- A young man wearing this shirt at a political rally sent a strong signal about his persona and ideology, one undetectable to the average citizen.

**A Unified Account**

- We do not want to say that a piece of clothing can bear propositional content.
- However, it is completely natural in sociolinguistics to say that clothing, hairstyles, and other aspects of presentation bear social meaning (a phenomenon Eckert calls *bricolage*).
- The existence of non-linguistic dogwhistles forces us to a non-propositional, social meaning account.
- This account can then be used for both linguistic and non-linguistic cases, providing a unified theory of the phenomenon.

# 18 Conclusion

**Chapter Summary**

- This chapter developed a comprehensive account of identifying dogwhistles.
- We showed how they emerge in RSA signaling games when messages are differentially linked with controversial personas across an audience.
- In such situations, it can be optimal for a speaker to use a dogwhistle to partition the audience and maximize their social utility.
- We also considered what kinds of linguistic and non-linguistic signals make for good dogwhistles.

**Looking Ahead**

- This chapter has focused primarily on the speaker's perspective: when is it rational to dogwhistle?
- The listener in this model only infers a persona and assigns an affective value.
- In the next chapters, we will take the listener's perspective.
- Chapter 5 will build an analysis of **enriching dogwhistles**—those that convey propositional content in addition to signaling a persona.

# Chapter 5: Enriching Dogwhistles

## 19 Introduction

**Recap: Identifying Dogwhistles**

- In the last chapter, we analyzed **identifying dogwhistles**.
- We treated them as signals that allow clued-in listeners to identify the speaker's social persona.
- Examples like Bush's "wonder-working power" or Stein's "big pharma" tag the speaker with a certain worldview or discourse familiarity.
- The analysis focused on how this identity recognition proceeds and affects listener attitudes toward the speaker.

**A New Type: Enriching Dogwhistles**

- This chapter turns to **enriching dogwhistles**.
- These do more than just identify a persona; they alter the semantic, truth-conditional content of the utterance that hosts them.
- For example, the dogwhistle "inner city" can be enriched from meaning "neighborhoods in built-up areas" to specifically mean "such neighborhoods inhabited primarily by people of color."
- Our aim is to propose an account for this type of dogwhistle.

**Chapter Roadmap**

1. Briefly review and critique previous accounts (Khoo's inferential view and our own prior work).
2. Propose a new account where recognition of a speaker's persona invites specific inferences.
3. Explicate the nature of personas and ideologies, and how they project beliefs and values.
4. Show how these projected beliefs ground the phenomenon of enriching dogwhistles.

## 20 Personas and Perspectives

**The Central Question**

- Our account so far has focused on social meaning, which was profitable for explaining the conventional-but-deniable nature of dogwhistles.
- However, this strategy falls short for enriching dogwhistles like "inner city."
- The question becomes: **How does the recognition of a persona contribute to, or induce, changes in truth-conditional meaning?**

### The Process of Enrichment

Imagine the following process for a savvy listener:

1. Hear someone use a dogwhistle (e.g., 'TRA' for 'trans rights activist').
2. Identify the dogwhistle and associate it with a relevant ideology (e.g., trans-exclusionary attitudes).
3. Draw conclusions about the beliefs the speaker holds, based on the tenets of that ideology.

We argue that enriching dogwhistles acquire their enriched meanings via this process.

### Two Types of Personas

The literature seems to discuss at least two types of personas:

1. **Social Affect Personas**:
   - Exemplified by the Obama BBQ case ("cool guy," "doofus").
   - Marks the speaker as having certain social feelings, likes/dislikes, and being open to certain kinds of discussion.
2. **Ideological/Political Personas**:
   - Exemplified by the "wonder-working power" case (evangelical).
   - Marks the speaker as subscribing to a particular ideology and brings in global assumptions about the world in the form of beliefs associated with that ideology.

### A Simple Formal Model of Ideology

- Our goal is not an exhaustive model of ideology, but one sufficient to show how enrichment works.
- The model requires two formal elements:
  1. A way to model how ideologies associate with approbation or disapprobation of various actions and persons.
  2. A way to model the association of personas with beliefs about the world.

### Formal Element 1: Affective Value ($\rho$)

- For the first element, we use a function, $\rho$ ('rate'), that assigns affective values.
- This function takes individuals (or kinds of individuals) as input and yields a positive or negative real number, representing approval or disapproval.
- **Example**: MAGA ideology might assign a high positive value to Donald Trump, while antifascist ideology assigns a high negative value.

$$\rho_{\text{MAGA}}(\text{Trump}) \gg 0$$

$$\rho_{\text{antifa}}(\text{Trump}) \ll 0$$

**From Predicates to Kinds: The 'cap' Operator**

- Ideologies often have attitudes toward groups or properties, not just specific individuals.
- We use the kind-mapping operator, $^\cap$, to turn predicates into individual-typed objects that $\rho$ can evaluate.
- This allows us to model attitudes toward abstract concepts.
- **Example**: A racist ideology can be modeled by giving a negative value to the kind 'Black person'.

$$\rho_{\text{racist}}(^\cap\text{black\_person}) < 0$$

**Formal Element 2: The Ideological Basis ($\mathcal{B}$)**

- Ideologies are also sets of beliefs about how the world is.
- These beliefs are modelable as a set of propositions, which we call the **basis** of an ideology, denoted $\mathcal{B}$.
- This ideological basis is the driver of enrichment for dogwhistles.
- **Examples of Basis Propositions**:
  - Anti-vaxx ideology: "Vaccines have negative effects."
  - Racist ideologies: Beliefs about the relative value and superiority of ethnic groups.
- These beliefs function to bridge gaps in reasoning.

**Formalizing Ideology**

- With these elements, we can formalize an ideology $\iota$ as a pair consisting of an affect-assigning function and an ideological basis:

$$\iota = \langle \rho, \mathcal{B} \rangle$$

- This structure provides the necessary machinery for our account of enrichment.

**Social Sincerity**

- For the analysis to work, we need to connect a speaker's signals to their actual beliefs.
- We propose a principle of **Social Sincerity**, an analogue of Gricean Quality for social meaning.

> **Social Sincerity**
> When a speaker signals an ideological persona, the basis of that persona must correlate with the speaker's actual beliefs. Speakers can only signal personas which have a basis in beliefs they actually have.

### Formalizing Social Sincerity

The principle requires that if a speaker's utterance is compatible with a persona, they must believe most of the propositions in that persona's ideological basis.

$$\forall s, u, \pi[\text{utter}(s)(u) \wedge \pi \in \mathbf{emf}(u) \wedge \iota_\pi \rightarrow$$
$$\text{MOST}(p \in \Pi_2(\iota_\pi))(\text{Bel}(s, p))]$$

*Note: This is a weak version. It could be strengthened with different quantifiers or by weighting beliefs by their "core-ness" to the ideology. For our current purposes, this is sufficient.*

### A Note on Communities of Practice

- How do personas become associated with ideologies in the first place?
- We think the answer lies in the idea of communities of practice.
- A dogwhistle relies on multiple communities using an expression in different ways, leading it to have multiple associated social meanings (personas).
- Once an expression has multiple personas associated with it, listeners who know about the practices in these communities can infer ideology from its use, assuming Social Sincerity.

## 21 Proposal: Personas Induce Enrichment

### The Flow of Interpretation

The enrichment process for an enriching dogwhistle works as follows:

1. The speaker produces an utterance with an enriching dogwhistle.
2. A savvy hearer recognizes the persona associated with the dogwhistle.
3. The hearer connects this persona to an associated ideology and its basis ($\mathcal{B}$).
4. A proposition in the basis interacts with the at-issue content of the utterance.
5. This interaction allows the computation of an inference, leading to enrichment.

### Application: "Inner City"

- Recall Paul Ryan's use of "inner city."
- A savvy listener recognizes this as signaling a quasi-racist persona.
- This persona is associated with a quasi-racist ideology, $\iota$.
- The basis of this ideology, $\Pi_2(\iota)$, contains the default proposition:

  $$\text{live\_inner\_city}(x) > \text{black\_person}(x)$$

  (People who live in the inner city are generally Black).
- This proposition combines with the at-issue content ("people in the inner city lack a culture of work") to derive the enriched, racialized inference.

**Comparison with Previous Views**

- This account is a clear improvement on previous work.
- **vs. general pragmatic enrichment**: Our prior work lacked a principled answer for how enrichment relates to particular dogwhistles. The new account links it directly to persona recognition.
- **vs. Khoo's inferential view**: Khoo's account struggled with coextensive expressions (why don't all synonyms for "inner city" work as dogwhistles?). Our account solves this: the inference is triggered by the persona associated with a *particular expression*, not just its general meaning.

**Revisiting the "Dred Scott" Dogwhistle**

- Previous accounts (Saul; Lo Guercio and Caso) analyzed Bush's reference to the *Dred Scott* decision as a Relevance implicature.
- **The problem**: It's savvy listeners (anti-abortion activists) who get the dogwhistle, but for them, the reference *is* relevant and needs no enrichment. It's naive listeners who would find it irrelevant, but they miss the dogwhistle.

**The "Dred Scott" Dogwhistle via Enrichment**

Our account provides a better explanation:

1. The term "Dred Scott" bears social meaning, signaling an anti-abortion persona to savvy listeners.
2. This persona is linked to an ideology with a rich propositional basis.
3. This basis includes the proposition:

   $$\Box\text{overturn(dred\_scott)} > \Box\text{overturn(roe\_vs\_wade)}$$

   (The necessity of overturning *Dred Scott* implies the necessity of overturning *Roe v. Wade*).
4. When Bush says he will appoint judges who would overturn *Dred Scott*, the enriched meaning (that they would also overturn *Roe*) follows as a direct entailment from the ideological basis.

No Gricean maxims are needed.

## 22 Other Evidence and Conclusions

**Other Evidence: Experiments on Bias**

- Experimental results on racial dogwhistles from Hurwitz and Peffley (2005) support our non-deflationary view.
- **The Finding**: After exposure to racial dogwhistles, White people were found to unconsciously assign stereotypes to racial minorities, but Black people were not, despite recognizing the dogwhistles.
- For White people, being pre-warned that the expression was a dogwhistle blocked this effect.

**Interpreting the Bias Experiments**

- This suggests that recognizing a dogwhistle and its associated persona does not automatically force the listener to make the enriching inference.
- Instead, the ideology associated with the persona can "bleed over" into a person's judgments, creating subconscious bias.
- However, participants who are already primed to resist the ideology (Black participants, or pre-warned White participants) are "inoculated" against this effect.
- What this means is that one can recognize the dogwhistle and what it means to do in terms of enrichment without actually performing the enrichment or adopting any elements of the ideology associated with the conveyed persona.
- This shows that a simple inferential account like Khoo's is too simple; the relationship between persona, ideology, and enrichment is more complex.

**Chapter Conclusion**

- This chapter provides a complete account of how enriching dogwhistles work.
- The main takeaway: Dogwhistles bear social meaning, not secondary truth-conditional content.
- This social meaning links to ideologies, which have propositional bases.
- Listeners who detect the dogwhistle can use the speaker's detected persona and its associated ideology to enrich the truth-conditional content of what was said.

**Looking Ahead**

- With this account in place, we can now extend it in various ways.
- The next two chapters will shift focus to the listener's perspective and the broader pragmatics of dogwhistles.
- We will explore what kinds of implicatures can be drawn over social meanings and how the use of dogwhistles should affect our assessment of a speaker's reliability.

# Chapter 6: Vigilance and Hypervigilance

## 23 Introduction: The Listener's Perspective

**Recap and Chapter Focus**

- The first half of this book established that persona construction is interactional.
- Speakers balance their desire for a persona with the audience's (dis)approval.
- We have focused on when it is optimal for a *speaker* to dogwhistle.
- This chapter now explores how *listeners* do and should react to dogwhistles.
- We will consider the behavior of the sociolinguistically aware listener, $L_1$.

**The Formal Answer**

- From a formal perspective, the answer to how a listener should reason is straightforward within the RSA framework.
- The optimal behavior for a sociolinguistically aware listener ($L_1$) is to reason about the speaker's ($S_1$) choice of message:

$$P_{L_1}(p|u) \propto P_{S_1}(u|p) \times P(p)$$

- The listener models how the speaker reasons about message choice relative to the audience.

**The More Interesting Question**

- Does this formal model successfully capture the *observed behavior* of listeners?
- We are at a slight disadvantage here, as the literature on dogwhistles has focused more on speakers and expressions than on rational listener behavior.
- This chapter begins to fill that gap by identifying two listener strategies:
  1. **Vigilance**
  2. **Hypervigilance**

## 24 Vigilance

**The Listener's Dilemma**

- When observing a potential dogwhistle, what is the proper reaction?
- The core problem is that the expression might be used innocently.
- A speaker may pick up a term from a political discourse without recognizing its dogwhistle quality.
- Even so, non-intentional dogwhistles can still have a pernicious effect on the discourse.

### The Role of Intent

- Our intuition is that the speaker's intent matters.
- If use is unintentional, we should be less likely to assign the speaker the taboo persona or the beliefs of the associated ideology.
- This means a simple "literal listener" model based on raw frequency is incorrect.
- Instead, we must consider if it's utility-maximizing for the speaker to be intentionally deceptive in the current context.
- Our bridge between intent and utility maximization is just that, on par, acts that are utility maximizing are evidence of intent.

### Vigilance Implicatures

- If it is profitable for the speaker to dogwhistle, a savvy listener should be *even more likely* to assign the taboo persona than raw frequency would suggest.
- This is a **vigilance implicature**.
- It is a type of implicature in the social meaning domain, analogous to those in the truth-conditional domain.
- A major result of this work is that our RSA-style account immediately predicts the existence of such implicatures.

### Recap: Implicatures in RSA Reference Games

- To understand vigilance implicatures, let's briefly revisit how standard implicatures work in RSA.
- We'll use the simple reference game from Frank and Goodman (2012).
- **Setup**: The speaker sees one of three objects (blue square, blue circle, green square) and must send a one-word message to a listener so they can identify it.

### The Reference Game: Literal Listener ($L_0$)

The literal listener assumes any object satisfying the predicate is equally likely. On hearing "blue," they have a 50/50 chance of picking the square or the circle.

Literal Listener ($L_0$)

| | blue square | blue circle | green square |
|---|---|---|---|
| *blue* | .5 | .5 | 0 |
| *green* | 0 | 0 | 1 |
| *square* | .5 | 0 | .5 |
| *circle* | 0 | 1 | 0 |

### The Reference Game: Pragmatic Speaker ($S_1$)

The pragmatic speaker reasons about $L_0$. When seeing the blue circle, she should prefer saying "circle" over "blue" because "circle" is perfectly informative, whereas "blue" is ambiguous.

Pragmatic Speaker ($S_1$)

| | blue square | blue circle | green square |
|---|---|---|---|
| *blue* | .5 | .33 | 0 |
| *green* | 0 | 0 | .66 |
| *square* | .5 | 0 | .33 |
| *circle* | 0 | .66 | 0 |

### The Reference Game: Pragmatic Listener ($L_1$)

The pragmatic listener reasons about the pragmatic speaker's choice.

- $L_1$ knows that $S_1$ would have likely said "circle" if she saw the blue circle.
- Therefore, hearing the less-specific message "blue" makes the blue *square* the more likely referent.

Pragmatic Listener ($L_1$)

| | blue square | blue circle | green square |
|---|---|---|---|
| *blue* | .6 | .4 | 0 |
| *green* | 0 | 0 | 1 |
| *square* | .6 | 0 | .4 |
| *circle* | 0 | 1 | 0 |

The probability bump from 0.5 to 0.6 for "blue square" is a pragmatic enrichment—an implicature.

### Vigilance Implicatures Formally

- The structure of our Social Meaning Games for dogwhistles is analogous to this reference game.
- A speaker considers the audience and reasons it is utility-maximizing to use a dogwhistle to signal a taboo persona to a sub-audience.
- A savvy listener, reasoning about this speaker, sees that dogwhistling is a likely move in this context.
- Therefore, the listener increases their likelihood of assigning the speaker the taboo persona associated with the dogwhistle.
- This is a vigilance implicature.

### The Jill Stein Example (Simplified)

To demonstrate this, we'll use a simplified version of the Jill Stein example.

- We assume an audience of uniformly savvy listeners who differ only in their affective values (preferences).

- To model the group, we can average the affective values across listeners.
- Let's assume an audience where anti-vaxxers outnumber pro-vaxxers 2-to-1, a situation where dogwhistling is rational.

### Model Setup

The idea is that "big pharma" is linked to anti-vax discourse in particular, but could possibly be used in anti-corporate speech.

- For the sake of simplicity, we take the phrase "corporate scientists" to fill the rest of the semantic space in this example, which idealization, but not a bad assumption.
- In particular, we see that listeners know it is not associated with anti-corporate, anti-vax speakers, because as savvy listeners they know these people use "big pharma".

**Likelihood of "big pharma" for savvy listeners**

| personas | $Pr(m\|p)$ |
|---|---|
| {pro-vaxxer, pro-corporate} | 0 |
| {pro-vaxxer, anti-corporate} | .5 |
| {anti-vaxxer, pro-corporate} | .75 |
| {anti-vaxxer, anti-corporate} | 1 |

**Likelihood of "corporate scientists" for savvy listeners**

| personas | $Pr(m\|p)$ |
|---|---|
| {pro-vaxxer, pro-corporate} | 1 |
| {pro-vaxxer, anti-corporate} | .5 |
| {anti-vaxxer, pro-corporate} | .25 |
| {anti-vaxxer, anti-corporate} | 0 |

We further assume that listeners have uniform priors over the space of possible personas for Stein.

**Priors for Stein's persona**

| | |
|---|---|
| {pro-vaxxer, pro-corporate} | .25 |
| {pro-vaxxer, anti-corporate} | .25 |
| {anti-vaxxer, pro-corporate} | .25 |
| {anti-vaxxer, anti-corporate} | .25 |

The final ingredient is a valuation function for each kind of listener.

- Once again, we are assuming that all listeners have a relative dislike for pro-corporate personas, but that antivaxxers also dislike pro-vaxx personas and pro-vaxxers dislike anti-vax personas.

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

| personas | Values |
|---|---|
| {pro-vaxxer, pro-corporate} | 1 |
| {pro-vaxxer, anti-corporate} | 2 |
| {anti-vaxxer, pro-corporate} | 19 |
| {anti-vaxxer, anti-corporate} | 20 |

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

| personas | Values |
|---|---|
| {pro-vaxxer, pro-corporate} | 19 |
| {pro-vaxxer, anti-corporate} | 20 |
| {anti-vaxxer, pro-corporate} | 1 |
| {anti-vaxxer, anti-corporate} | 2 |

From these two types of listeners we can construct a listener that represents the average of the audience.

- We are considering a situation where it is rational to dogwhistle, so let's assume that the anti-vaxxers outnumber the pro-vaxxers 2 to 1.

**$\nu_{\mathbf{L}}(\mathbf{p})$ for the average savvy listener**

| personas | Values |
|---|---|
| {pro-vaxxer, pro-corporate} | 7 |
| {pro-vaxxer, anti-corporate} | 8 |
| {anti-vaxxer, pro-corporate} | 13 |
| {anti-vaxxer, anti-corporate} | 14 |

We can now compute the effects for our three kinds of interlocutors. The literal listener $L_0$, in a by now familiar way, just rebalances after ruling out those personas that are simply inconsistent with the message.

**Literal Listener**

| | +vax,+corp | +vax,-corp | -vax,+corp | -vax,-corp |
|---|---|---|---|---|
| *corporate scientists* | .57 | .29 | .14 | 0 |
| *big pharma* | 0 | .22 | .33 | .44 |

The speaker, who is pragmatically aware, will use messages at a certain frequency aimed at maximizing utility given the speaker's preferences for certain personas.

- Given the particular cost structure we imposed (and temperature 1), what we see is that pragmatically aware speakers will be preferring *big pharma* for all anti-vax personas, even pro-corporate ones, and *corporate scientists* for all pro-vax personas, even anti-corporate ones.

**Speaker**

| | +vax,+corp | +vax,-corp | -vax,+corp | -vax,-corp |
|---|---|---|---|---|
| *corporate scientists* | 1 | .68 | .03 | 0 |
| *big pharma* | 0 | .32 | .97 | 1 |

The reason is that a pragmatically aware listener will know that a pragmatically aware speaker will be using *big pharma* to signal anti-vax personas at a higher rate than we might otherwise think because it is utility maximizing for the literal listener.

- This means that the pragmatic listener should assign anti-vax personas to speakers that use *big pharma* at a rate greater than we should expect given the literal social meaning of the phrase.

- This is exactly what we see. Compare the following to the chart for the literal listener.

**Pragmatic Listener**

| | +vax,+corp | +vax,-corp | -vax,+corp | -vax,-corp |
|---|---|---|---|---|
| *corporate scientists* | .58 | .4 | .02 | 0 |
| *big pharma* | 0 | .14 | .42 | .44 |

In particular, the overall probability mass assigned to anti-vax personas shifts from .77 to .86, and the source of uncertainty—i.e., "big pharma" could be being used to signal anti-corporate leanings—is removed.

- This is the vigilance implicature. The listener who knows the speaker maximizes utility by dogwhistling is more likely to dogwhistle, and so should be more likely to detect the whistle.

### Vigilance and Intention

- The model also captures the role of intention, a distinction first made by Saul (2018).
- We can use legal reasoning as an analogy: motive (*utility* in our model) is not the same as intent (*mens rea*), but it is relevant evidence.
- Vigilance implicatures model this reasoning: by seeing that the speaker has a motive (a utility payoff) to dogwhistle, the listener finds it easier to believe they intended to do so.

### Modeling Unintentional Dogwhistles

- While intent is not an explicit part of the model, we can account for it by adjusting priors.
- **Example**: A newscaster reporting on a politician's dogwhistle.
- A listener should have a low prior that the newscaster intends to project the taboo persona.
- With this low prior, the listener is unlikely to tag the newscaster with the persona, though the possibility is not eliminated. The use is not blameless.
- The model also captures the "amplifying" effect: even if the newscaster is not blamed, the original dogwhistle's message is spread to a wider audience.

## 25 Hypervigilance

### Introducing Hypervigilance

- Vigilance focuses on the speaker's interests. Listeners infer a dogwhistle is more likely when it's utility-maximizing for the speaker.
- **Hypervigilance** emerges from a more direct comparison of speaker and listener interests.
- It's a phenomenon often seen on social media, where people are quick to label expressions as dogwhistles.

**Example: Keir Starmer and "Family"**

- Keir Starmer, leader of the UK Labour Party, is seen as being on the right side of his party.
- This has put him at odds with the left flank that supported his predecessor, Jeremy Corbyn.
- As a result, online sleuths on the left comb through his pamphlets and tweets, looking for conservative dogwhistles.
- His use of the word "family" has been seized upon as an anti-LGBTQ dogwhistle.
- **Example Tweet Reply**: "...a Lab leader going pastel blue tradfam dogwhistle...".

**The Dynamic of Hypervigilance**

- This is likely a case of hypervigilance.
- Because Starmer is seen as an ideological opponent (or at least not a true ally), his co-partisans are hypervigilant and motivated to hunt for examples.
- In contrast, Jeremy Corbyn uses the word "family" without similar accusations, because his leftist persona is seen as authentic.
- The key factor is the perceived ideological distance between the speaker and the listener.

**Modeling Hypervigilance**

- This is difficult to model in the standard RSA framework, as listeners don't label expressions as dogwhistles; they infer personas.
- The locus of hypervigilance must be in the likelihood parameter, $P(m|p)$.
- Our proposal: hypervigilance involves a shift in this parameter based on perceived ideological distance.

**The Hypervigilant Listener**

The hypervigilant listener operates in two steps:

1. **Become Skeptical**: No longer trust their understanding of the speaker's social meanings. They back off from their learned values for $P(m|p)$.
2. **Become Paranoid**: Re-calculate the likelihood based on ideological distance.

We can connect this to work on credibility in cheap talk games by Farrell (1993): a novel message is only credible to the extent that sender and receiver interests align. Hypervigilant listeners perceive misaligned interests.

### The Hypervigilant Model

- We model the alignment of interests as the ideological distance, $D$, between the speaker's persona and the listener's persona.
- The hypervigilant listener uses this distance to determine the likelihood of a message.

$$\text{HypervigilantL}_0(p|m) \propto ([m](p) \times D(p_{L_0}, p)) \times P(p)$$

- In this state, the probability that a speaker with an enemy persona used a certain message is proportional to how distant (i.e., how "enemy") that ideology is.
- Any expression consistent with an enemy ideology becomes evidence for it.

### Hypervigilance and Discourse Breakdown

- This model is intentionally perverse, because the phenomenon itself is perverse from a cooperative standpoint.
- The speaker must "walk on eggshells"; any message that could possibly be construed as negative will be read as negatively as possible.
- This often leads to discourse breakdown, where the speaker's best move is not to play at all.
- The goal is no longer communication, but rooting out linguistic evidence that the speaker bears a disliked persona.

### Is Hypervigilance Always Bad?

- This behavior is not always the listener's fault. Intentional dogwhistling is itself a bad faith, duplicitous act.
- One should not be a good faith listener to a bad faith speaker; ending the conversation can be a defensible move.
- Furthermore, moderately hypervigilant listeners may play an important role in the broader community.
- By aggressively flagging potential dogwhistles, they can act as an "early warning system," helping the population detect new or subtle signals, even if there are many false positives.

## 26 Conclusions

### Chapter Conclusion

- This chapter has shifted focus to the listener, exploring how one should react to a possible dogwhistle.
- We introduced **vigilance implicatures**, which arise naturally from our RSA model when considering pragmatically aware listeners. Listeners become more likely to detect a dogwhistle when it's a profitable move for the speaker.
- We also modeled **hypervigilance**, a more aggressive and adversarial strategy where listeners use perceived ideological distance to interpret messages.
- This accounts for the conversational breakdowns often seen in online political discourse.

**Looking Ahead**

- The next chapter will continue this focus on listener reactions to dogwhistles.
- We will shift from the adversarial case of detecting and punishing dogwhistles to the role of social meaning in building trust.
- We will see that trust can be built, even if an utterance is not strictly truthful, as long as the social meaning signal is strong enough.

# Chapter 7: Dogwhistles and Trust

## 27 Introduction: The Puzzle of Trust

### Trust in Linguistics

- Questions of trust are often discussed in linguistics, though not always by that name.
- We are advised to provisionally believe what people say because we assume they follow the Gricean Maxim of Quality (don't lie).
- In Chapter 5, we proposed a different notion, *Social Sincerity*: we provisionally assume people sincerely project personas linked to beliefs they actually hold.
- Both notions rest on the idea that speakers are sincere in what they convey.

### The Puzzle of Trusting the Untruthful

- A problem arises when certain speakers are trusted despite being known to "play fast and loose with the facts".
- This is common in political discourse.
- **Example**: Donald Trump. Journalists have devoted immense time to exposing his false claims, but his supporters seem to trust him nonetheless.
- On standard epistemological views, this is a surprise. Given that someone is consistently untruthful, why should we ever trust them?

### Chapter Goals and Roadmap

This chapter uses our tools to address this puzzle in three parts:

1. Spell out the puzzle using a formal theory of source evaluation and reliability.
2. Propose a theory of how ideological considerations are valued alongside truth-conditional content.
3. Extend the theory of reliability to cover ideologically-based trust, grounded in expectations about what a speaker with a certain ideology will do.

### The Central Claim

- The solution to the puzzle lies in how people value different aspects of communication.
- We claim that people trust messy testimonial agents like Trump because, in certain contexts, they value shared ideological beliefs more than they value true-speaking.
- This can be a surprisingly rational mode of behavior for political settings.

# 28 Evaluating Information Sources

### A Formal Model of Reliability

- To make the puzzle more precise, we use the theory of reliability from McCready (2015).
- This theory addresses how one can determine whether content from a particular source should be believed.
- The theory has two parts:
  1. A method for assigning initial judgments of reliability.
  2. A method for updating those judgments based on interaction.

### Initial Judgments: Heuristics

- An agent first uses a set of heuristics to determine if a source has properties associated with reliability.
- Properties taken into account include the source's profession, personal presentation, gender, race, etc.
- Some of these are genuinely useful (e.g., trusting a linguist on linguistics), while others lead to epistemic injustice (e.g., distrusting someone based on their gender or race).

### Updating Judgments: Interaction History

- The initial judgment is only a starting point, modified by interaction with the source.
- Each discourse move by a source contributes to a history.
- These moves are modeled as tuples $\langle \varphi, V \rangle$, where $\varphi$ is the content and $V$ is its truth value (T, F, ?).
- The reliability of an agent, $R_a$, is the frequency of their T-valued moves:

$$R_a = \frac{t_a}{t_a + f_a}$$

- Each truth-tracking move raises perceived reliability, while each failure to track truth lowers it.

### From Reliability to Belief

- The model uses a dynamic semantics where an agent tracks multiple information sources in separate submodels, or "information states" ($\sigma_i$).
- These states are ranked by a plausibility ordering, $\preceq_a$, based on the reliability of each source.

$$i \prec_a j \Longleftrightarrow Rel_a(i) < Rel_a(j)$$

- To determine global belief, all information states are merged. In case of conflict, information from the more reliable source overrides the less reliable one.

### The Puzzle Restated

- In this formal model, an agent who is consistently non-truth-tracking will have their reliability rating fall with each false statement.
- As their rank on the $\prec_a$ ordering drops, the content they provide is less likely to survive the merge into the global belief state.
- Thus, the content provided by a known liar should not be believed.
- **The Puzzle**: Why does this fail for politicians like Trump, who are still trusted by their supporters despite a long history of false statements?
- **The Answer**: The model is missing a crucial component—social meaning and ideology.

## 29 Ideology and Trust

### Recap: The Dogwhistle Model

- Recall from previous chapters that dogwhistles send information about a speaker's persona and ideology.
- We modeled this with Social Meaning Games where utility depends on the recovery of a persona and the listener's evaluation of it.
- The key insight is that messages package both social/ideological content and truth-conditional content.

### The Basis of Affective Value: Homophily

- What is the basis for assigning positive or negative value ($\nu$) to a persona?
- We propose a simple metric: **similarity**, based on the principle of *homophily*.
- Homophily: Common values make social interaction rewarding.
- The intuition is "I like people who are like me," or more formally, "I like people who have social personas similar to mine".
- This principle is widely used and forms part of the basis for algorithms driving social media bubbles.

### A New Weighted Utility Function

- We can now solve the puzzle by letting hearers determine trust through a mix of truth-conditional reliability and social meaning, weighted differently in different contexts.
- We follow Henderson and McCready (2018) and weight the two components of utility with values $\delta$ and $\gamma$.

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

- $\delta$ indexes the value placed on social meaning.
- $\gamma$ indexes the value placed on truth-conditional meaning.

**Exploring the Weights**

- Setting $\delta = 0$ and $\gamma > 0$ models a style of communication where social meaning is disregarded, like "science" traditionally construed.
- Setting $\gamma = 0$ and $\delta > 0$ gives a "post-truth" style of communication, where facts are irrelevant and only social persona matters.
- **The Trump Case**: When $\delta \gg \gamma$ (the value of social meaning vastly outweighs the value of truth), we get Trump-voter-style trust.

**Believed vs. Trusted**

- In a system where social meaning is highly valued, "reliability" can increase based on social signaling alone, even if truth-tracking is poor.
- In this context, "reliability" is better understood as a high degree of persona similarity that induces homophily.
- This makes it possible to judge an individual unreliable on facts, but still trust them because you believe they share your goals and interests.
- The idea is that an agent with a similar persona can be **trusted**, without precisely being **believed**.

**Formalizing Social Trust**

- We can extend McCready's model to track social trust over time.
- Game iterations now include the projected persona: $\langle \varphi, V, \pi \rangle$.
- Initial trust is just the similarity between the speaker's first projected persona and the interpreter's own persona:
$$trust_a^1 = sim(\pi_1, P)$$
- Trust is then updated by averaging with the similarity of the currently expressed persona:
$$trust_a^{i+1} = \frac{sim(\pi_i, P) + trust_a^i}{2}$$

**When is this Rational?**

- A strategy that devalues truth might seem aberrant, but it can be rational in certain scenarios.
- In politics, where voters may not know much about policy, a politician's ideology (signaled via social meaning) can be a more useful heuristic for making voting decisions than their specific factual claims.
- A sophisticated listener might set $\delta \gg \gamma$ early in a conversation to figure out a speaker's ideology, then switch to $\gamma \gg \delta$ later to monitor their truthfulness once their persona is known.

# 30 Social Hedges: Fake News and Fig Leaves

**Social Hedges**

- We now turn to communicative moves used by speakers who rely on a social-trust-based strategy.
- **Social Hedges**: Moves that manipulate the perceived importance of social vs. truth-conditional meaning.
  1. "Take them seriously, not literally."
  2. Crying "fake news."
  3. Discursive fig leaves.

**Move 1: "Seriously, Not Literally"**

- This phrase has been used by many pundits to describe Trump's communication style.
- What does it mean? In our framework, the meaning is clear.
- It is an explicit exhortation to set the weights such that $\delta \gg \gamma$, or even to set $\gamma = 0$.
- We are instructed not to pay attention to the literal truth of the utterances, but to take the speaker's social persona (as a fighter, a businessman, etc.) seriously.

**Move 2: "Fake News"**

- Calling a statement "fake news" is a signal to remove it from truth-relevant consideration.
- It functions as a denial operator combined with a pragmatic effect.

  **The Fake News Operator**
  $FN(\varphi) =$
  1. $Deny(\varphi)$
  2. $\neg Low(Sim(\pi_r, \pi_h))$ (Asserts that the original speaker's persona, $\pi_r$, is dissimilar to the hearer's, $\pi_h$)
- This move destroys trust from both the truth-conditional and social meaning perspectives.
- It strategically forces listeners into the social meaning game by making the truth-conditional side of the calculation irrelevant.

**Move 3: Fig Leaves**

- A fig leaf is a bit of content "tagged onto" an utterance to thinly veil an objectionable attitude.
- **Examples**:
  - "I'm not racist, but..." followed by a racist statement.
  - Trump's claim that Mexican immigrants are "rapists," followed by "And some, I assume, are good people".
- These are hedges that aim to deflect the inference of a particular social persona (e.g., a racist one).

### Modeling Fig Leaves

- We can view fig leaves as a request to reset the utility weighting parameters for a specific utterance.
- A fig leaf hedge for social meaning is a request to set $\delta = 0$ for that utterance.
- If the hedge is accepted, the hearer will not engage in persona recovery for that utterance and will only consider its truth-conditional content.
- Of course, the hearer can reject the hedge and compute the speaker as not just racist, but also disingenuous for trying to hide it.

## 31 Conclusions

### Chapter Conclusions

- This chapter has shown how trust can be separated from truth-conditional reliability.
- We can trust the testimonially unreliable if we value social similarity over truth-tracking.
- This can be a rational strategy in contexts like politics, where a candidate's ideology may be more important to a voter than their day-to-day factual accuracy.

### A Prescriptive Consequence

- A consequence of this analysis is that fact-checking a politician who relies on a social-trust strategy is unlikely to sway their supporters.
- Those supporters are already playing a game where truth is devalued (a low $\gamma$).
- To be effective, oppositional discourse must challenge the politician's social persona, either by showing it is insincere or that its associated ideology is harmful to the listener themselves.

### Prognosis for Political Discourse

- This analysis helps explain why exposing the lies of some populist politicians has little effect on their base of support.
- If persona projection is primary, it becomes difficult to hold politicians accountable, as their ideologies are signaled rather than stated directly.
- This makes the state of political discourse look muddy and opens the possibility of politics being a noncooperative context where the interests of politicians and voters are opposed.
- While this book cannot solve this problem, we hope it helps to clarify its structure.

# Chapter 8: Beyond Dogwhistles

## 32 Summary of the Book

**Summary: What We've Done**

- This book has provided an analysis of dogwhistles, situating them in the broader pragmatic landscape and giving an account in terms of persona projection.
- We divided dogwhistles into two types:
  - **Identifying dogwhistles**: Indicate the speaker's (political) views and persona.
  - **Enriching dogwhistles**: Also alter the literal meaning of the sentence.
- We showed that dogwhistles occupy a liminal space between conventional and nonconventional meaning, where previous accounts have failed.

**Summary: The Core Theory**

- Chapters 3-5 steered a middle course, using tools from Rational Speech Act (RSA) theory and the formalization of sociolinguistic identity.
- **Chapter 4 (Identifying Dogwhistles)**: Modeled as coded messages that allow clued-in listeners to infer the speaker's social persona.
- **Chapter 5 (Enriching Dogwhistles)**: Showed that enrichment relies on persona identification, which activates an associated ideology and allows for further inferences.

**Summary: Listener Pragmatics and Trust**

- **Chapter 6 (Vigilance)**: Extended the RSA model to sociolinguistically aware listeners, who work harder to draw conclusions about personas. We also modeled hypervigilance as a function of ideological distance.
- **Chapter 7 (Trust)**: Addressed the puzzle of why people trust politicians who consistently ignore the truth. We proposed a theory where trust can be based on ideological similarity rather than just truth-tracking.

**This Chapter's Roadmap**

This final chapter considers extensions and consequences of the theory:

- **Section 8.2**: "Mask-off" moments, where speakers choose *not* to dogwhistle.
- **Section 8.3**: Situating social meanings within broader pragmatic theory.
- **Section 8.4**: The relationship between personas, perspectives, and standpoint epistemology.
- **Section 8.5**: The scope of the model beyond dogwhistles to other forms of coded speech.

# 33 Mask-Off Moments and Discursive Health

### Mask-Off Moments

- Our theory so far explains why it's rational to dogwhistle in a mixed audience.
- But it doesn't address the opposite phenomenon: Why would a speaker choose to abandon dogwhistles and instead make an overt, extreme appeal?
- We call these situations **"mask-off" moments**.

### A Typology of Unmasking

We propose that mask-off moments come from three general sources, which are likely to emerge under political polarization:

1. Changes in the speaker's beliefs about dogwhistle **effectiveness**.
2. Changes in the speaker's beliefs about their **audience**.
3. Changes in the **value** the speaker assigns to presenting a certain persona.

### Factor 1: Perceived Ineffectiveness

- A speaker might stop dogwhistling if they believe it's no longer effective.
- This can happen if the speaker believes the audience *already* thinks they hold the taboo persona, regardless of what they say.
- This often occurs after being "called out" for more subtle signals, which leads to increased scrutiny and hypervigilance from opponents.
- **The slogan**: "If I'm already canceled, I'll just speak my mind."
- The speaker concludes it's best to focus on appealing to those who already approve of their views.

### Factor 2: Increased Value of Persona

- A speaker might also stop dogwhistling if they radically increase the affective value ($\nu_S$) they assign to the persona associated with the dogwhistle.
- As $\nu_{S_1}(p)$ for some persona $p$ tends to $\infty$, dogwhistling becomes non-optimal.
- It is better to make an overt appeal and ensure all audience members assign you $p$, even if they don't like $p$.
- The speaker's own affective value for $p$ will swamp whatever the audience values.
- **The slogan**: "I don't care what you think of me if you don't think like me."
- This represents a move toward overt extremism and polarization.

**Factor 3: Change in Perceived Audience**

- A speaker may stop dogwhistling if they come to believe they are only addressing "same-believers".
- If the audience is no longer perceived as mixed, there is no incentive to use a dogwhistle.
- This happens easily on social media, where one can't see the full audience and may pay more attention to "likes" (from allies) than to combative comments.
- **The slogan**: "I'm not talking to you anymore."

**A Counterintuitive Conclusion**

- Dogwhistling is usually seen as deceitful and uncooperative.
- However, given the current state of the world, we argue that we should be more concerned when people *stop* dogwhistling.
- If speakers dogwhistle, it means they still care what people with different views think and believe dialogue is still possible. They think they're not "canceled" yet.
- Ceasing to dogwhistle means they have given up on everyone outside their bubble.
- In this sense, the use of dogwhistles can be a sign of a *relatively* healthy discursive environment.

## 34 Lessons on Social Meaning

**Broader Questions about Social Meaning**

- Our analysis of dogwhistles speaks to broader questions about the nature of social meaning.
- What kinds of interactions exist between social and truth-conditional meanings?
- Do social meanings come in different types, like other non-truth-conditional content?
- What is the structure of the social meaning domain?

**Types of Social Meaning**

- Our analysis suggests at least two types of social meanings modeled via personas.
    1. **Non-ideological**: Signals qualities like friendliness or approachability (e.g., Obama's use of *cookin'*).
    2. **Ideological**: Signals a social identity like "racist" or belief in certain propositions.
- These two types are clearly different and require distinct formal treatments.

**Interaction with Truth-Conditional Content**

- Social meanings clearly interact with truth-conditional content, at least for the ideological type.
- **Case 1**: Persona recognition affects interpretation. If I think you are projecting a racist persona, I am more likely to interpret your subsequent ambiguous utterances as having racist content.
- **Case 2**: Persona recognition affects trust. Observing a dogwhistle for an ideology I oppose will lower my trust in the speaker. If I know someone holds beliefs I deem irrational (e.g., flat-earthers), I am less likely to believe their factual claims.

**Structure of the Social Meaning Domain**

- For ideologies, the notion of an ideological "core" (the intersection of beliefs held by all subscribers) imposes some structure.
- This structure can be viewed as inducing a Voronoi tessellation, where ideologies are partitioned based on similarity to a core set of propositions.

**Do Gricean Maxims Apply to Social Meaning?**

- **Quality**: Has an analogue in our "Social Sincerity" principle.
- **Relevance**: Is vague enough to apply, but doesn't seem to do much extra work.
- **Manner**: Is opaque even in the truth-conditional context, so it's hard to draw conclusions.
- **Quantity**: Seems inapplicable. It is very difficult to find pairs of social-meaning-bearing expressions where one cleanly entails the other in the required way.
- **Maximize Presupposition**: Does not seem to apply. We don't have conscious control over all social signals (like accent), and requiring people to always reveal their ideology is unreasonable.

# 35 Standpoints

**Connecting to Standpoint Epistemology**

- Our theory can be used to make formal sense of the idea of a **standpoint**.
- **Standpoint Epistemology**: The idea that one's position in society can grant access to unique knowledge that others might not have, provided one is attentive to that positioning.
- A standpoint is usually cashed out in terms of identities like gender, race, sexuality, etc.

**A Formal Proposal for Standpoints**

- We propose thinking of a standpoint as a pair:
  1. A particular kind of probability distribution (a "good" prior).
  2. The capacity to *sincerely* project particular personas associated with an identity.

- The "consciousness" required by standpoint theory can be modeled by the agent's ability to be sincere about their persona.

**The Problem of Sincerity and Identity**

- How can one be "sincere" about an identity persona like "woman"? We can't use an ideological basis, as that would be essentialist.
- We propose a different notion of sincerity: an agent can sincerely project an identity persona if they take themselves to be of the identity that persona signals.
- **Example**: I can project a 'woman' persona if the predicate 'woman' truly applies to me: $woman(C_a) = 1$.

**The Need for Uptake**

- A worry arises: Is subjective self-identification enough? (e.g., the Rachel Dolezal case).
- No. To occupy a standpoint, it's not enough to sincerely project a persona; one's interlocutors must also **accept** it.
- This is because the epistemic advantage of a standpoint comes from being *treated* as a member of that group and learning from those experiences.
- Thus, standpoints are relational and require hearer uptake.

**A Presuppositional Link**

- A final lesson from this analysis is a dependency between meaning domains.
- The ability to project certain personas presupposes certain truth-conditional facts about identity.
- For instance, to present with a 'dyke' persona, one must be able to sincerely self-predicate 'woman'.
- This shows that truth-conditional meaning and social meaning are not fully independent, at least at the pragmatic level of sincere use.

# 36 The Scope of the Theory

**Are Dogwhistles the End of the Story?**

- To us, it seems clear that they aren't.
- The theory speaks quite generally to coded communications that depend on recognition of social identity.
- **Example**: The "hanky code" used in gay communities to covertly signal sexual interests. These aren't dogwhistles, but our theory could apply to them.

**All Speech is Coded Speech**

- Once the scope of social coding is recognized, one might wonder if any expression lacks a social meaning.
- Even the most basic word choice (e.g., ‘dog‘ vs. ‘canine‘) carries social meaning about register, formality, and community.
- From this perspective, every term becomes a dogwhistle, because every term carries social meaning that yields different information to different audiences.

**Conclusion**

- The theory we have developed for the narrow analysis of dogwhistles has a much more general application.
- It is fundamentally a theory about how differential probabilities (sociolinguistics) lead to differential meanings (formal pragmatics).
- We hope this book helps make sense of how these two fields interact and opens up new possibilities for the future.

# Appendix

Wither logarithms (Shannon 1948):

# A Mathematical Theory of Communication

By C. E. SHANNON

## INTRODUCTION

THE recent development of various methods of modulation such as PCM and PPM which exchange bandwidth for signal-to-noise ratio has intensified the interest in a general theory of communication. A basis for such a theory is contained in the important papers of Nyquist[1] and Hartley[2] on this subject. In the present paper we will extend the theory to include a number of new factors, in particular the effect of noise in the channel, and the savings possible due to the statistical structure of the original message and due to the nature of the final destination of the information.

The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have *meaning*; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem. The significant aspect is that the actual message is one *selected from a set* of possible messages. The system must be designed to operate for each possible selection, not just the one which will actually be chosen since this is unknown at the time of design.

If the number of messages in the set is finite then this number or any monotonic function of this number can be regarded as a measure of the information produced when one message is chosen from the set, all choices being equally likely. As was pointed out by Hartley the most natural choice is the logarithmic function. Although this definition must be generalized considerably when we consider the influence of the statistics of the message and when we have a continuous range of messages, we will in all cases use an essentially logarithmic measure.

The logarithmic measure is more convenient for various reasons:

1. It is practically more useful. Parameters of engineering importance such as time, bandwidth, number of relays, etc., tend to vary linearly with the logarithm of the number of possibilities. For example, adding one relay to a group doubles the number of possible states of the relays. It adds 1 to the base 2 logarithm of this number. Doubling the time roughly squares the number of possible messages, or doubles the logarithm, etc.

2. It is nearer to our intuitive feeling as to the proper measure. This is closely related to (1) since we intuitively measures entities by linear comparison with common standards. One feels, for example, that two punched cards should have twice the capacity of one for information storage, and two identical channels twice the capacity of one for transmitting information.

3. It is mathematically more suitable. Many of the limiting operations are simple in terms of the logarithm but would require clumsy restatement in terms of the number of possibilities.

The choice of a logarithmic base corresponds to the choice of a unit for measuring information. If the base 2 is used the resulting units may be called binary digits, or more briefly *bits,* a word suggested by J. W. Tukey. A device with two stable positions, such as a relay or a flip-flop circuit, can store one bit of information. $N$ such devices can store $N$ bits, since the total number of possible states is $2^N$ and $\log_2 2^N = N$. If the base 10 is used the units may be called decimal digits. Since

$$\begin{aligned}\log_2 M &= \log_{10} M / \log_{10} 2 \\ &= 3.32 \log_{10} M,\end{aligned}$$

[1]Nyquist, H., "Certain Factors Affecting Telegraph Speed," *Bell System Technical Journal,* April 1924, p. 324; "Certain Topics in Telegraph Transmission Theory," *A.I.E.E. Trans.,* v. 47, April 1928, p. 617.

[2]Hartley, R. V. L., "Transmission of Information," *Bell System Technical Journal,* July 1928, p. 535.

## References


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

Burnett, Heather. 2017. Sociolinguistic interaction and identity construction: The view from game-theoretic pragmatics. *Journal of Sociolinguistics* .

Burnett, Heather. 2019. Signalling games, sociolinguistic variation and the construction of style. *Linguistics and Philosophy* .

Eckert, Penelope. 2008. Variation and the indexical field. *Journal of sociolinguistics* 12:453–476.

Farrell, Joseph. 1993. Meaning and credibility in cheap-talk games. *Games and Economic Behavior* 5:514–31.

Frank, Michael C, and Noah D Goodman. 2012. Predicting pragmatic reasoning in language games. *Science* 336:998–998.

Henderson, Robert, and Elin McCready. 2018. Dog-whistles and the at-issue/not-at-issue distinction. In *Secondary content*, ed. Daniel Gutzmann and Katherine Turgay, 191–210. Brill.

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. *Public Opinion Quarterly* 69:99–112.

Khoo, Justin. 2017. Code words in political discourse. *Philosophical Topics* .

Lo Guercio, Nicolás, and Ramiro Caso. 2022. An account of overt intentional dogwhistling. *Synthese* 200:1–32.

McCready, Elin. 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.

Podesva, Robert. 2004. On constructing social meaning with stop release bursts. In *Sociolinguistics Symposium*, volume 15, 1–5.

Podesva, Robert J. 2007. Phonation type as a stylistic variable: The use of falsetto in constructing a persona 1. *Journal of sociolinguistics* 11:478–504.

Saul, Jennifer. 2018. Dogwhistles, political manipulation, and philosophy of language. *New Works on Speech Acts* 360–383.

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