# Lexical Racialization Examined through Machine Learning

Kelly E. Wright
LSA Department of Linguistics

Information Analysis and Retrieval
Winter Seminar Series
23 April 2019

# Outline

- Background on the Social
- Background on the Linguistic
- Dataset and Methods
- Interpretation
- Questions and Discussion
- Analysis Wall
  - Options for visualization going forward

# Measuring Implicit Human Biases Through the Statistical Properties of Language

# Racialization

- Happens when a word with no preexisting racial connotation comes to describe people of color asymmetrically.
  - *Thug*
- A new Polysemous Sense
  - One word, Multiple meanings

# *Articulate*

## More than a microaggression

- Alim and Smitherman (2012) *Articulate While Black*
  - Critical discussion begins with President Obama in 2008

# *Articulate*

More than a microaggression

- Alim and Smitherman (2012) *Articulate While Black*
  - Critical discussion begins with President Obama in 2008
- In the US, standard linguistic varieties are ideologically associated with Whiteness and formal education
  - Calling a minority articulate notes surprisal at standard language usage in the absence of these traits.
  - Typically followed up with, he went to school at X.

# Racism

- Evolving concept
  - 1970s shift from overt to covert
  - Housing as an example, de facto segregation to de jure segregation (redlining, highway project, etc.)
- New type of prejudice
  - Colorblindness; Post-Racial Society (Coates 2012; Henry 2010; Hill 2008; Bonilla-Silva 2000)

# Racism

- Evolving concept
  - 1970s shift from overt to covert
  - Housing as an example, de facto segregation to de jure segregation (redlining, highway project, etc.)
- New type of prejudice
  - Colorblindness; Post-Racial Society (Coates 2012; Henry 2010; Hill 2008; Bonilla-Silva 2000)
- Not acquired through explicit lessons
  - “Part of the individual’s rational ordering or her perceptions of the world” (Lawrence 1995)

# Racism

- How do we see social ideology reflected in text?
  - (aka language use)
- Wright (2017) hypothesis:
  - at the lexical level.

# Lee Atwater

You start out in 1954 by saying, “[N-Word, N-word, N-Word].” By 1968 you can't say “[N-Word]” — that hurts you. Backfires. So you say stuff like **forced busing**, **states' rights** and all that stuff. You're getting so abstract now [that] you're talking about cutting taxes, and all these things you're talking about are totally economic things and a byproduct of them is [that] blacks get hurt worse than whites… I'm saying that if it **is getting that abstract, and that coded, that we are doing away with the racial problem** one way or the other. You follow me — because obviously sitting around saying, “We want to cut this,” is much more abstract than even the busing thing, and a hell of a lot more abstract than “[N-word, N-word].”

# Covert Racism

*Bussing* in our above example

- Racialization happens
  - Regular words, with no racial connotation, become asymmetrically applied to the descriptions of actions or events of actors of color.

# Covert Racism

- Racialization happens
- Not inherently negative

# The Myth of Unbeatable Black Athleticism

- Folk ideology, mistakenly applied throughout US history
- Black athletes are *not* exceptional or naturally suited for physical activity, or for violent displays or prowess in team or individual sport.

# Wright (2017)

- Black athletes are described predictably differently
  - Exceptionality and Animalistic traits
- White athletes
  - Leadership or Skill-based terms

# Caliskan et al. 2017

## Latent Semantic Analysis

- Able to recreate bias results from Implicit Association Tasks with Word Vector models

# Caliskan et al

N= Number of Subjects  $N_T$= Target Words  $N_A$= Attribute Words  D = Effects Sizes  P = Pvalues

| Target words | Attribute words | Original finding | | | | Our finding | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Ref. | *N* | *d* | *P* | $N_T$ | $N_A$ | *d* | *P* |
| Flowers vs. insects | Pleasant vs. unpleasant | (5) | 32 | 1.35 | $10^{-8}$ | 25 × 2 | 25 × 2 | 1.50 | $10^{-7}$ |
| Instruments vs. weapons | Pleasant vs. unpleasant | (5) | 32 | 1.66 | $10^{-10}$ | 25 × 2 | 25 × 2 | 1.53 | $10^{-7}$ |
| European-American vs. African-American names | Pleasant vs. unpleasant | (5) | 26 | 1.17 | $10^{-5}$ | 32 × 2 | 25 × 2 | 1.41 | $10^{-8}$ |

- Word Vectors created from Target word stimuli and Attribute word stimuli
  - Semantic Field; Collocation Frequency
  - Google Corpus

# Caliskan et al. 2017

- Word Vector models from Google Corpus
- Latent Semantic Analysis
  - Related vectors to census data

# Caliskan et al. 2017

- Veridical data by way of statistics gathered from census data
- Output from LSA analysis
- Recreated Biases seen in the IATs

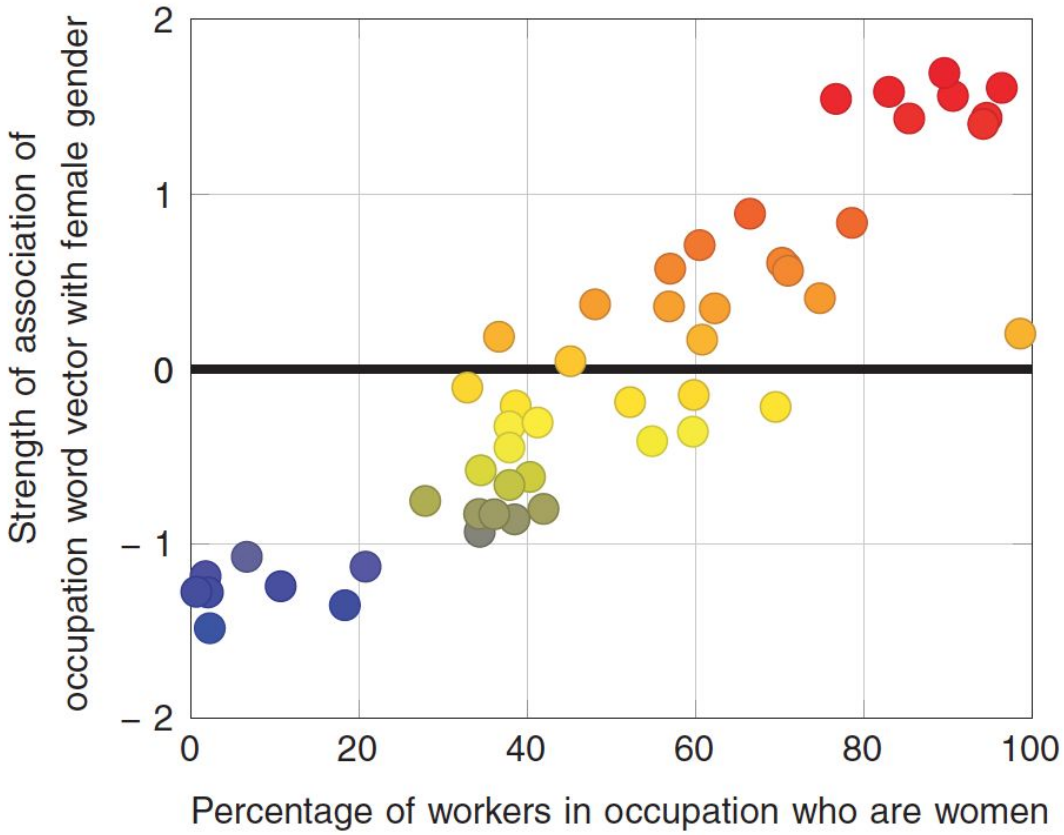



**Fig. 1. Occupation-gender association.** Pearson's correlation coefficient $\rho = 0.90$ with $P < 10^{-18}$.

# Caliskan et al. 2017

- Issue a call to action that my research is heading.

The number, variety, and substantive importance of our results raise the possibility **that all implicit human biases are reflected in the statistical properties of language**. Further research is needed to test this hypothesis and to compare language with other modalities, especially the visual, to see if they have similarly strong explanatory power. Our results also suggest a null hypothesis for explaining **origins of prejudicial behavior in humans**, namely, the implicit **transmission of ingroup/outgroup identity** information through language. That is, before providing an explicit or institutional explanation for why individuals make prejudiced decisions, one must show that it was not a simple **outcome of unthinking reproduction of statistical regularities absorbed with language**. Similarly, before positing complex models for how stereotyped attitudes perpetuate from one generation to the next or from one group to another, **we must check whether simply learning language is sufficient to explain (some of) the observed transmission of prejudice**."

What are there ways in which lexical semantic space is renegotiated as new (racialized) meaning is bound to old (non-racialized) words?

# Lexical

- Words (and lemmas)
- Mental Lexicon
- Shapes, Pronunciations, Meanings, Rules of Deployment
  - Knowing the meaning of a lexical items means you know the semantic entailments it brings with it

# Formal Semantics

## The Polysemy Problem

- The simplest processes of semantic change have proven reliably difficult to model.
- Spam: potted meat → Spam: junk email
  - Both meanings active, how do we choose?

# Trier’s Semantic Field Theory

- Trier’s Semantic Field Theory (1932)
- In Polysemous relationships, meaning is epiphenomenal
  - Not from inherent qualities of the lexeme (spam always is potted meat), but rather from its context or application (but when we use it is this way it means junk email).

# Field Semantics

- Levels of Categorization and Contrast
- Semantic Field Theory
  - Trier (1932)
  - a group of words with interrelated meanings can be categorized under a larger conceptual domain.
  - You can know *red* without also knowing *scarlet*

# Field Semantics

- Issue of polysemy
  - What do we do when we have two meanings for the same starting point. What fields needs to be accessed and active for us to semantically process a word?
- How is meaning organized cognitively and applied materially?

# Field Semantics

- Knowledge of Semantic Entailment is acquired
  - So too with ideology
- Lexical Knowledge and Social Ideology are Intimately Bound
  - Seeking to define the link

# Dataset

RSEAC

- **R**acialized **SE**mantics in **A**thletics **C**orpus
  - 120 Athletes
  - 60 White; 60 Black
  - 30 Male; 30 Female
  - 15,500 lexemes
  - 8.5 million total words

# RSEAC

- Combines previous methodologies
  - Metadata behind Sports Journalism
  - Highly controlled journalistic frame.
  - Longform and event pieces, describing individuals, not teams
- Variation based on patterns of actual disparities baked into the data, making racialization isolable.

# Collection Methods

- Search Engine Optimization Search
  - Search Engine Results Pages
- Advanced Search
  - “Serena Williams”
- Cleaning and concatenation bash script
  - Brings down text per article, removes noise
  - Saves as Subcorpora

# Machine Learning

- Support Vector Machine
  - Counting Stuff
- Random Forest Modeling Task
  - Analyzing Counted Stuff

# Support Vector Machine

- Learning Algorithm, trained to predict athlete race from lexical token counts
  - Subcorpora organized by athlete

# Lexical Usage Asymmetries

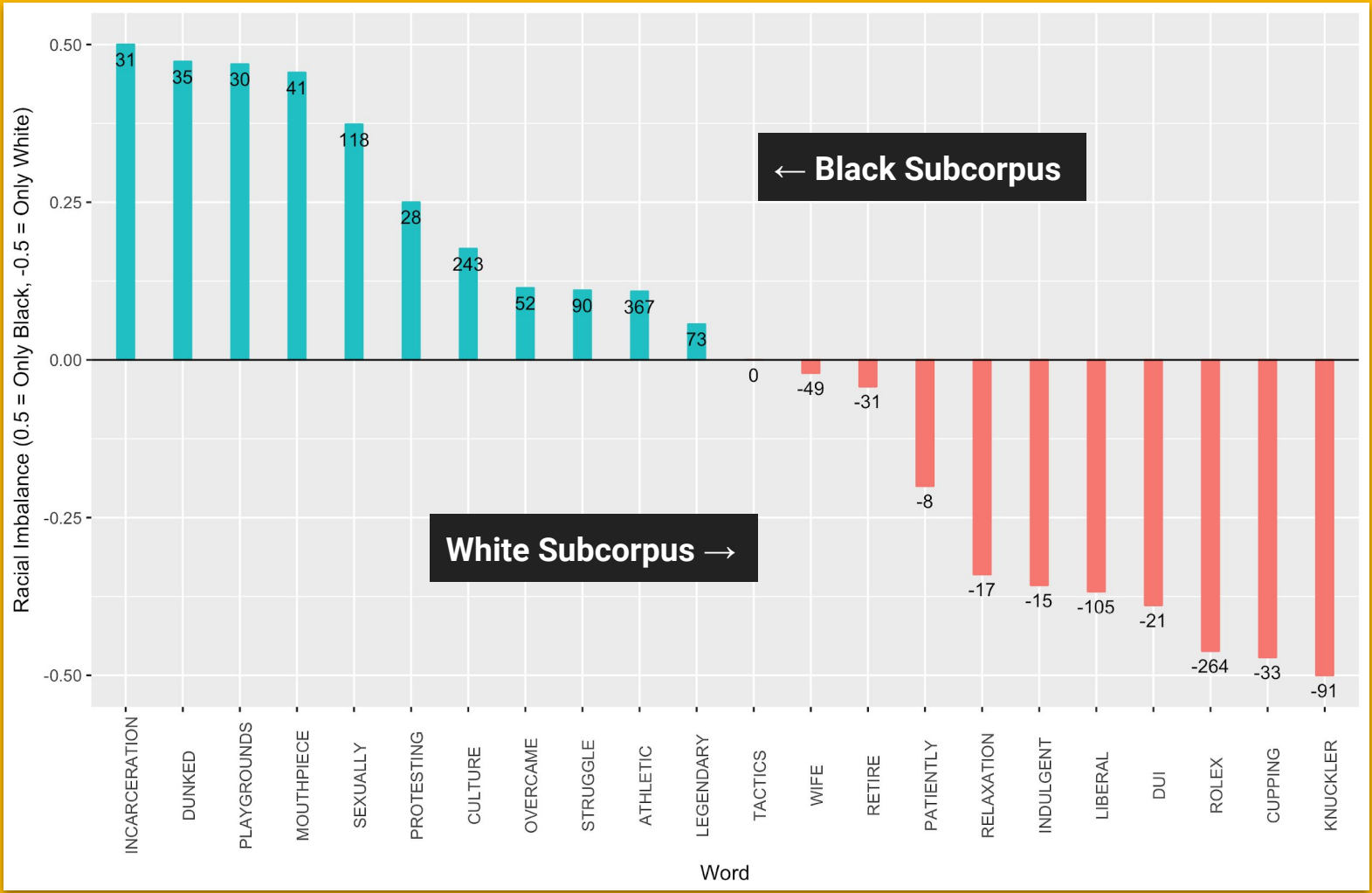
*Culture*

- Culture occurs 2:1 ratio
- Black athletes are discussed as
  - Infusing their own culture into the sport
  - Altering the sport's culture with their ethnic presence

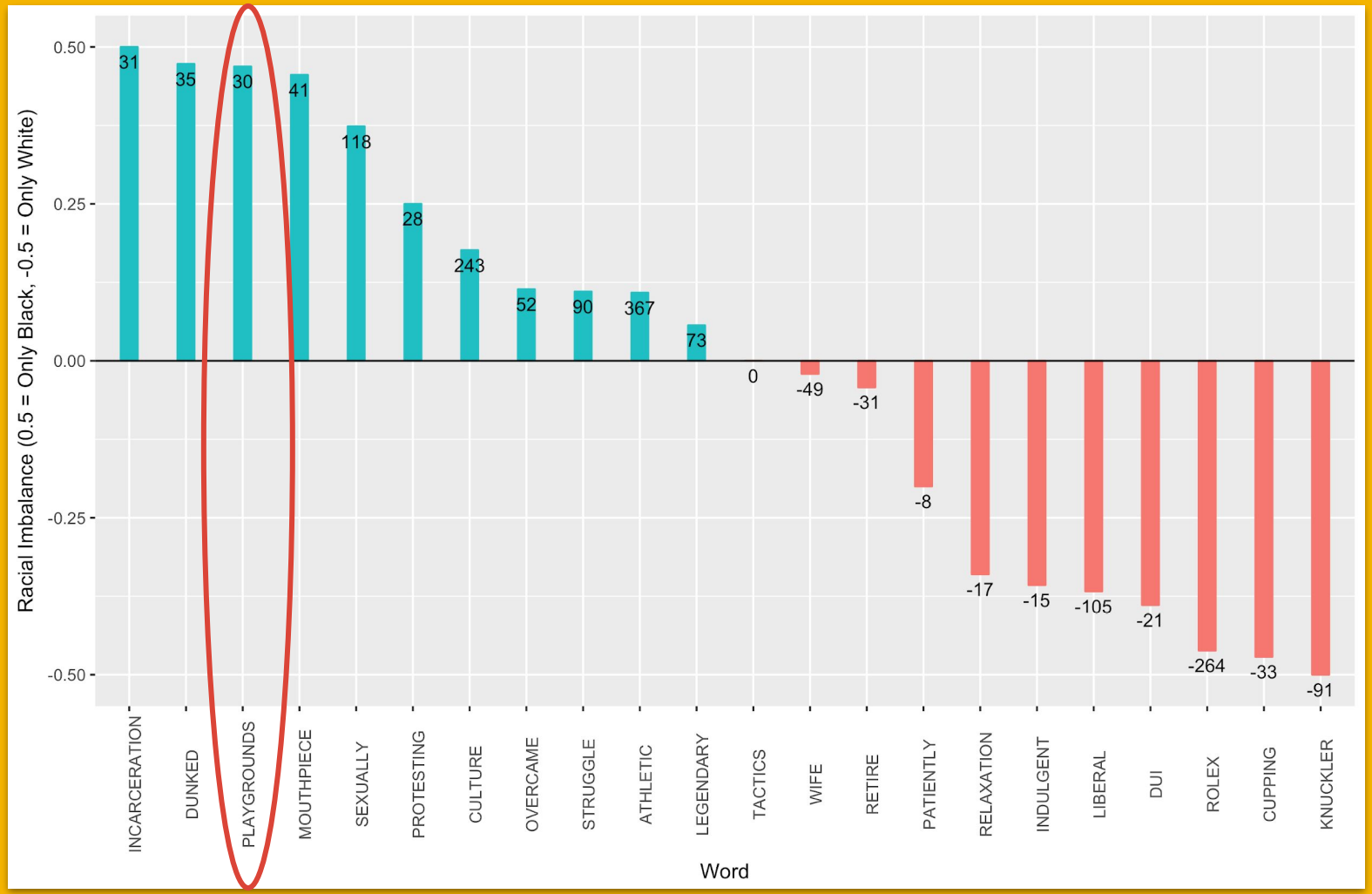
# Lexical Usage Asymmetries

*Culture*

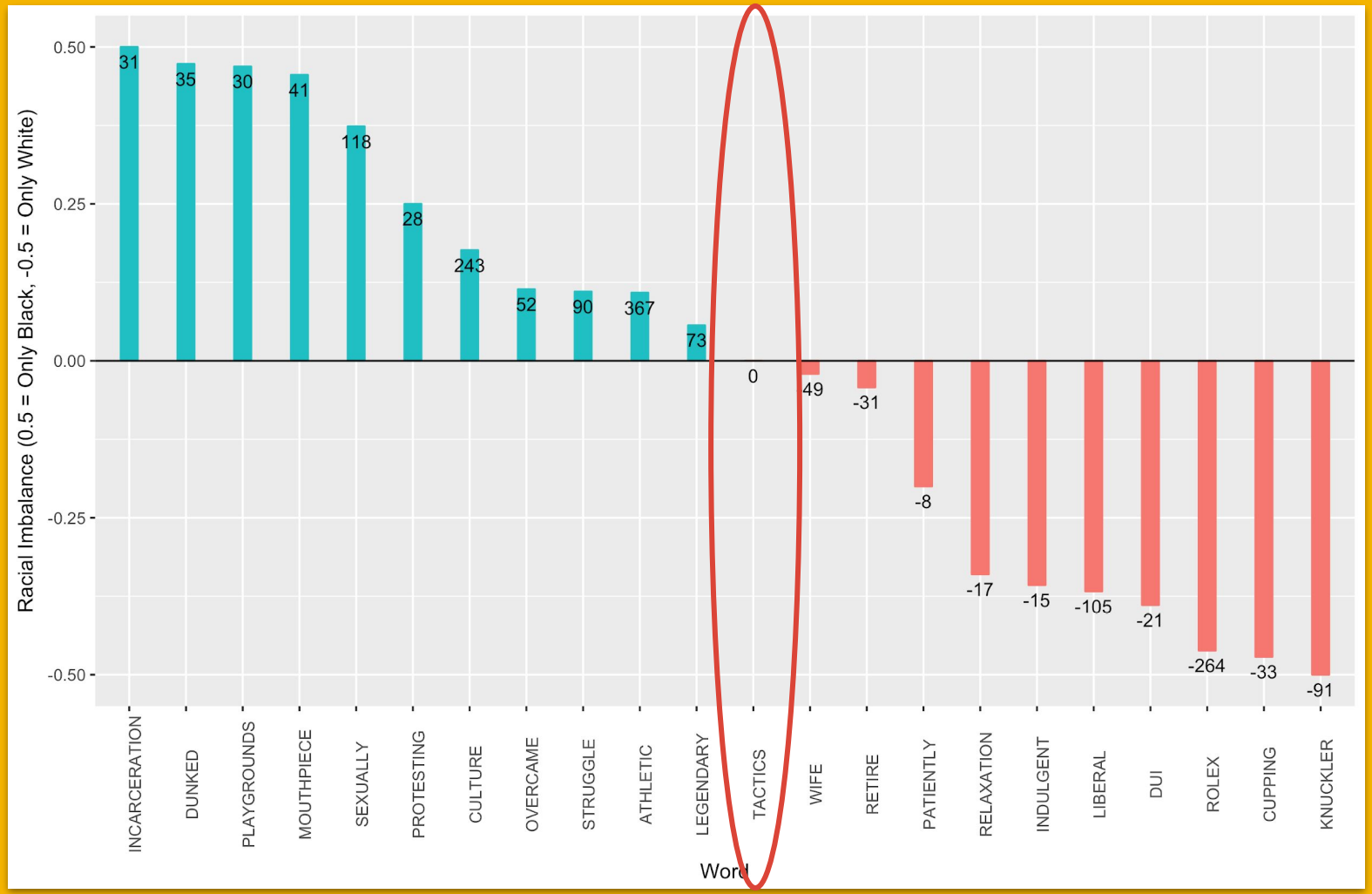
- Culture occurs 2:1 ratio
- Black athletes are discussed as
  - Infusing their own culture into the sport
  - Altering the sport's culture with their ethnic presence
- Culture occurrences in the White subcorpus
  - The culture of the sport itself, not White culture as such
- Two very different lexical sense of *culture*, the usage of which is determined by the race of the referent.


← Black Subcorpus
White Subcorpus →
Racial Imbalance (0.5 = Only Black, -0.5 = Only White)
0.50
0.25
0.00
-0.25
-0.50
31
35
30
41
118
28
243
52
90
367
73
0
-49
-31
-8
-17
-15
-105
-21
-264
-33
-91
INCARCERATION
DUNKED
PLAYGROUNDS
MOUTHPIECE
SEXUALLY
PROTESTING
CULTURE
OVERCAME
STRUGGLE
ATHLETIC
LEGENDARY
TACTICS
WIFE
RETIRE
PATIENTLY
RELAXATION
INDULGENT
LIBERAL
DUI
ROLEX
CUPPING
KNUCKLER
Word

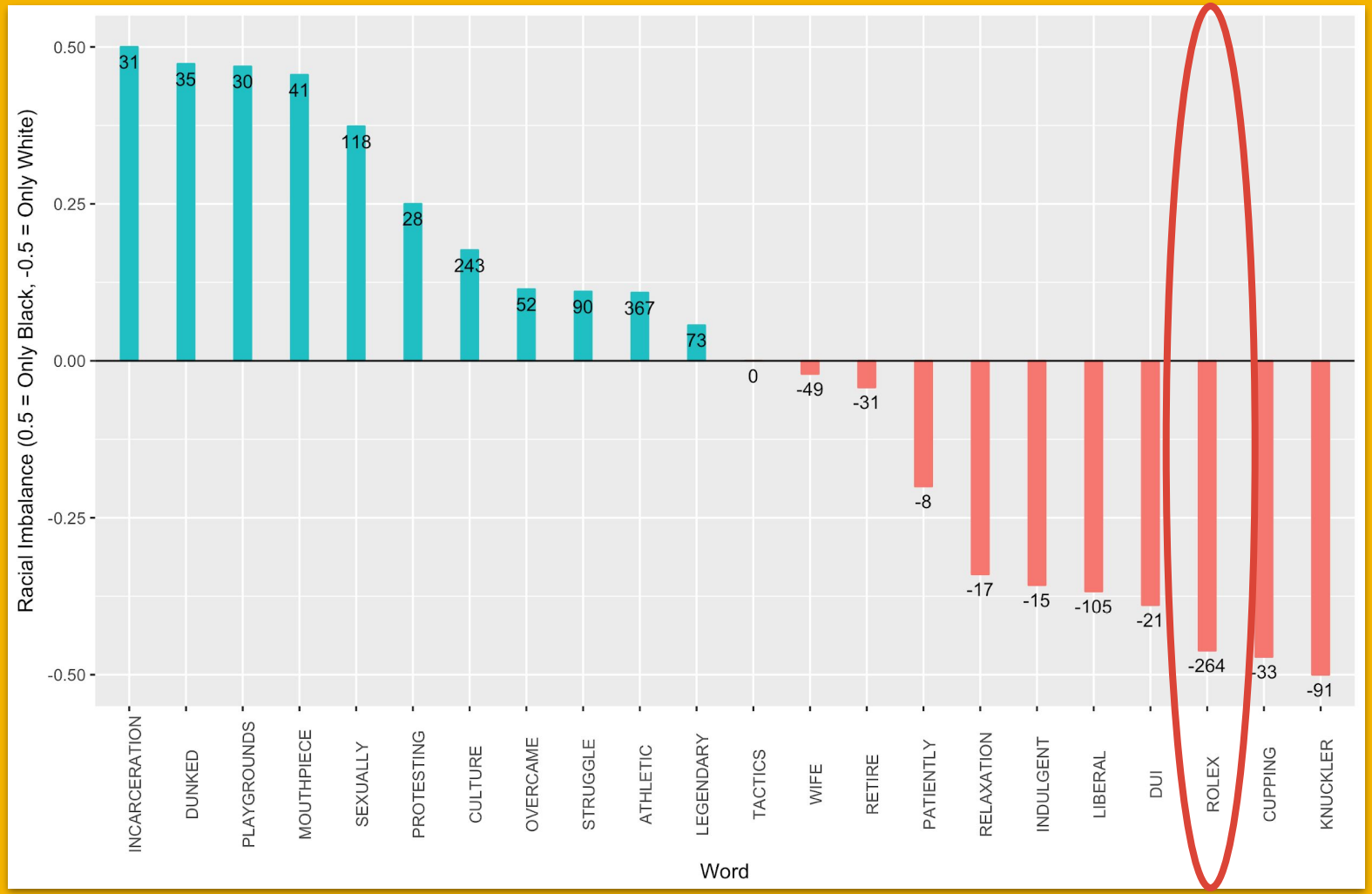
**Playgrounds**
31:1 or
97% In Black
Subcorpus



**Tactics**
43:43
Balanced across subcorpora



**Rolex**
11:275 or
96% in White
Subcorpus



| Athlete | Race Probability |
|---|---|
| Eric Berry | 3% |
| Chamique Holdsclaw | 3% |
| Brittney Griner | 3% |
| Anthony Davis | 3% |
| Jackie Joyner Kersee | 3% |
| Alia Atkinson | 3% |
| Hope Solo | 96% |
| Phil Mickelson | 96% |
| Andrew Luck | 96% |
| Ronda Rousey | 96% |
| Drew Brees | 96% |
| Katie Ledecky | 96% |

- The SVM sorted Racial Probability by Athlete Subcorpus
- The most impressive result here is a lack of gradience in the probabilities of category membership.

# Random Forest Modeling

- A learning algorithm building multiple decision trees and modeling based on the most accurate.
- Trained on the same dataset

# Random Forest

- Predict class membership by building decision trees
- See a portion of the data, use it to predict class membership on the rest of the data.
- 90/10 training ratio.

| Lexical Type | Importance | Black Sum | White Sum |
|---|---|---|---|
| BLOCKS | 0.2092 | 520 | 102 |
| NIKE | 0.1515 | 660 | 217 |
| COAST | 0.1511 | 132 | 214 |
| EFFORTS | 0.1451 | 310 | 205 |
| AVERAGED | 0.1421 | 573 | 109 |
| ATHLETIC | 0.1130 | 1029 | 662 |
| WONDERFUL | 0.1075 | 204 | 379 |
| UNDEFEATED | 0.1064 | 241 | 74 |
| WHOM | 0.1052 | 363 | 227 |
| APPEARED | 0.0958 | 545 | 333 |
| SOCIAL | 0.0904 | 872 | 560 |
| LLC | 0.0871 | 104 | 222 |

- RandomForest lets us crack into Lexical Importance to Categorization
- Model outputs predict which words were most useful in the racial categorization task.

# Implications

- Supports Caliskan et al. (2017) and Garg et al (2017)
- Supports the Distributional Hypothesis
- Supports Trier’s assertion that meaning is epiphenomenal
- Algorithms are super racist (Speer 2017)

# Future Directions

- Complete Vector Analysis
  - Lund and Burgess (1995)
  - Caliskan et al. (2017)
- BETTER VISUALIZATION!!!
  - (I have ideas; I need help)
- Extend time depth and sport balance in RSEAC
  - Apparent time studies
- Expand and test on new datasets
- Henderson & McCready (2018)
  - Dogwhistles
  - Where world knowledge and self knowledge mesh to control semantic judgements

# Thanks!

Kelly E. Wright

@raciolinguistic

kellywri@umich.edu