Actors and Events: Racialized Semantic Intensity Over Time

Kelly E. Wright
University of Michigan
LSA Department of Linguistics

NARNiHS @ The KFLC
Research Incubator
17 April 2020

Outline

These slides are available at

- Lexical Racialization
- Background and Previous Research
- Machine Learning Analysis
- Apparent Time (ongoing work!)
- Questions Going Forward

Kernel Idea

Let's Incubate!

- Seeking HS applications and methods to develop my research questions
- While my dataset covers 108 years, I have yet to view my current results in an HS frame.
- I need your help!

To what extent is the racialization of individual lexical items affected by the notoriety of the actors under discussion, or the (social/world) events surrounding their performance?

To Date

Side project, Schmide project

- Evidence of racialization: Semantic Intensity (2017)
- Isolated individual lexical items (2019)
- Prepped data for more informative visualization (2020)

Lexical Racialization

- Happens when a word with **no preexisting racial connotation** comes to describe people of color asymmetrically.
  - *Thug*

Lexical Racialization

- Happens when a word with **no preexisting racial connotation** comes to describe people of color asymmetrically.
  - *Thug*
- Doesn’t mean they are *never* applied to White people, significantly less often.

Field Semantics

- Trier (1932) Semantic Field Theory
  - 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

Re Racialization

- Semantic Field Theory
- Superordinate concept is *Blackness*
  - A group of words with interrelated meanings are used to describe Black actors, events, and spaces.
- Lexical Knowledge and Social Ideology are Intimately Bound
  - Seeking to define the link

Racism

- Not acquired through explicit lessons
  - “Part of the individual’s rational ordering of her perceptions of the world” (Lawrence 1995)
- Structural
  - We don’t have to be active participants to have our cognitive schema affected.

Ideological Change

Reflected in Lexical, Semantic Change

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

The Myth of Unbeatable Black Athleticism

- **Folk ideology**, mistakenly applied throughout US history

The Myth of Unbeatable Black Athleticism

- **Folk ideology**, mistakenly applied throughout US history
- Black athletes are *believed to be* exceptional or naturally suited for physical activity, or for violent displays or prowess in team or individual sport.

The Myth of Unbeatable Black Athleticism

Busted!

- Black athletes are NO DIFFERENT than White athletes in terms of capabilities for physical activity or for (violent) displays or prowess in team or individual sport.

Dana Mastro

(2011)

- Athletes described differently by race

Dana Mastro

(2011)

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

Kelly Wright

(2017)

- Racialization at the lexical level
- Serena Williams and Semantic Intensity
  - Avoidance of the "and she's Black" construction, because colorblindness.
    - The G.O.A.T at 21 years old.

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
- 108 year time depth

RSEAC

- Composed of longform and event pieces, describing individuals, not teams
- Highly controlled journalistic frame.
- Metadata behind Sports Journalism

Machine Learning

A script in R

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

Machine Learning

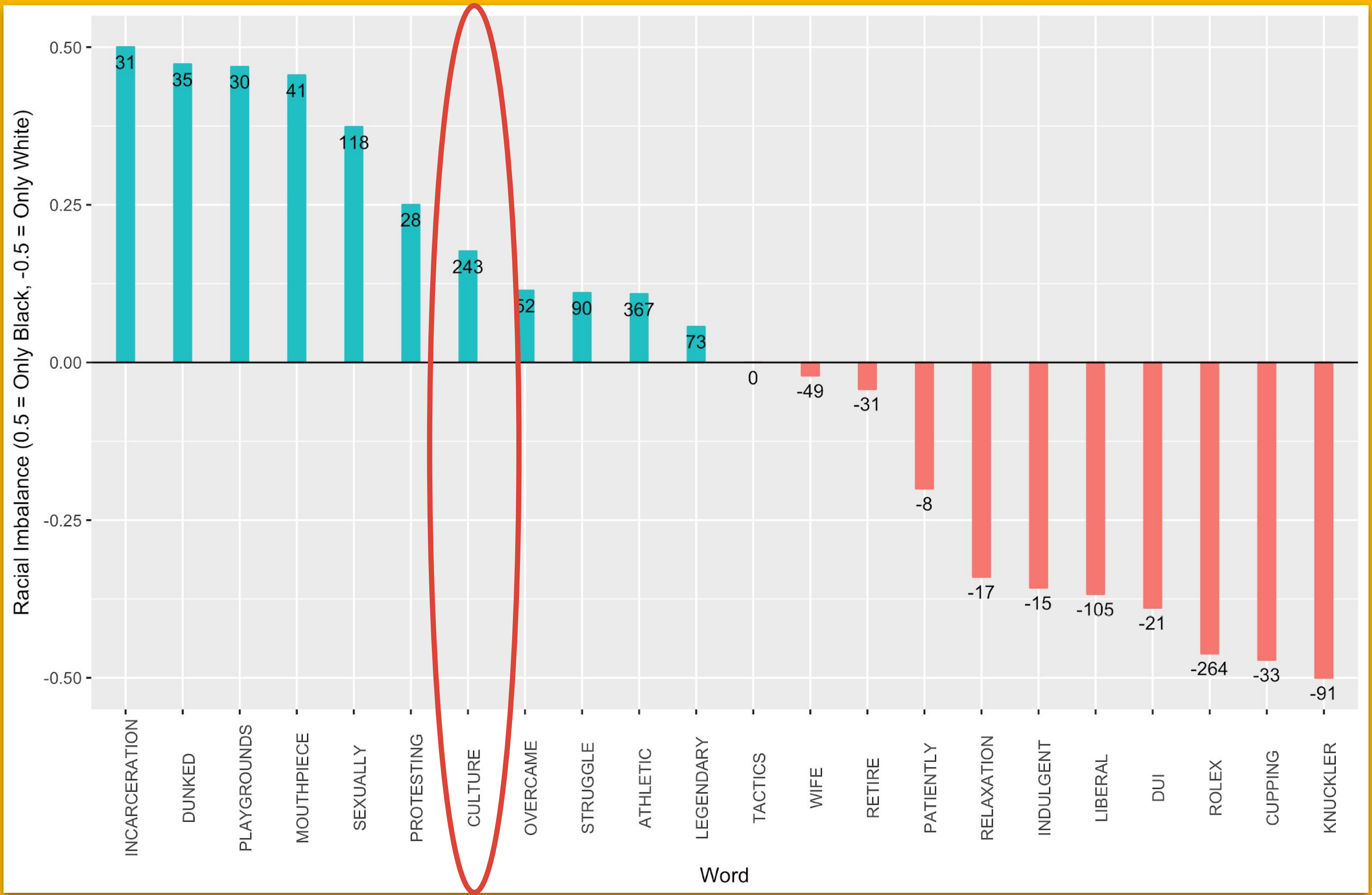
Step One
Counting Stuff

- A Support Vector Machine is a learning algorithm
  - I trained my SVM to predict athlete race with lexical token counts for each athlete as input.
  - Categorization task

Culture
2:1 ratio

Black Subcorpus →

White Subcorpus →



Lexical Usage Asymmetries

Culture

- ❖ Culture occurs 2:1 ratio
  - ➢ Twice as often in the Black subcorpus

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

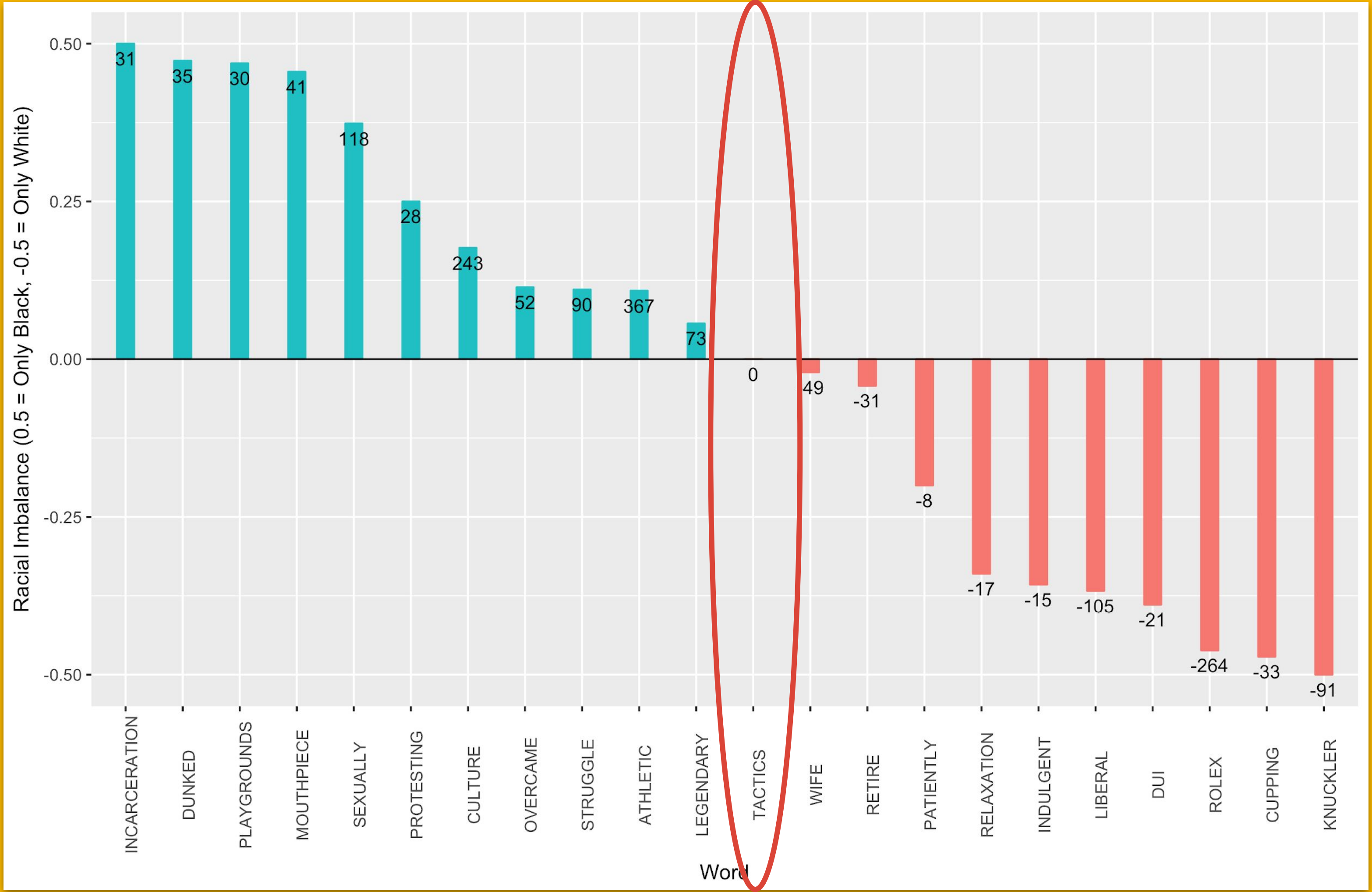
- Culture occurs 2:1 ratio
- Black athletes are discussed as
  - Infusing their own culture into the sport
  - Altering the sport’s culture with their ethnic presence
- *Culture* occurrences in the White subcorpus
  - The culture of the sport itself, not White culture as such
  - Not in reference to the athlete

Tactics
43:43

Black Subcorpus →

Balanced across subcorpora

White Subcorpus →

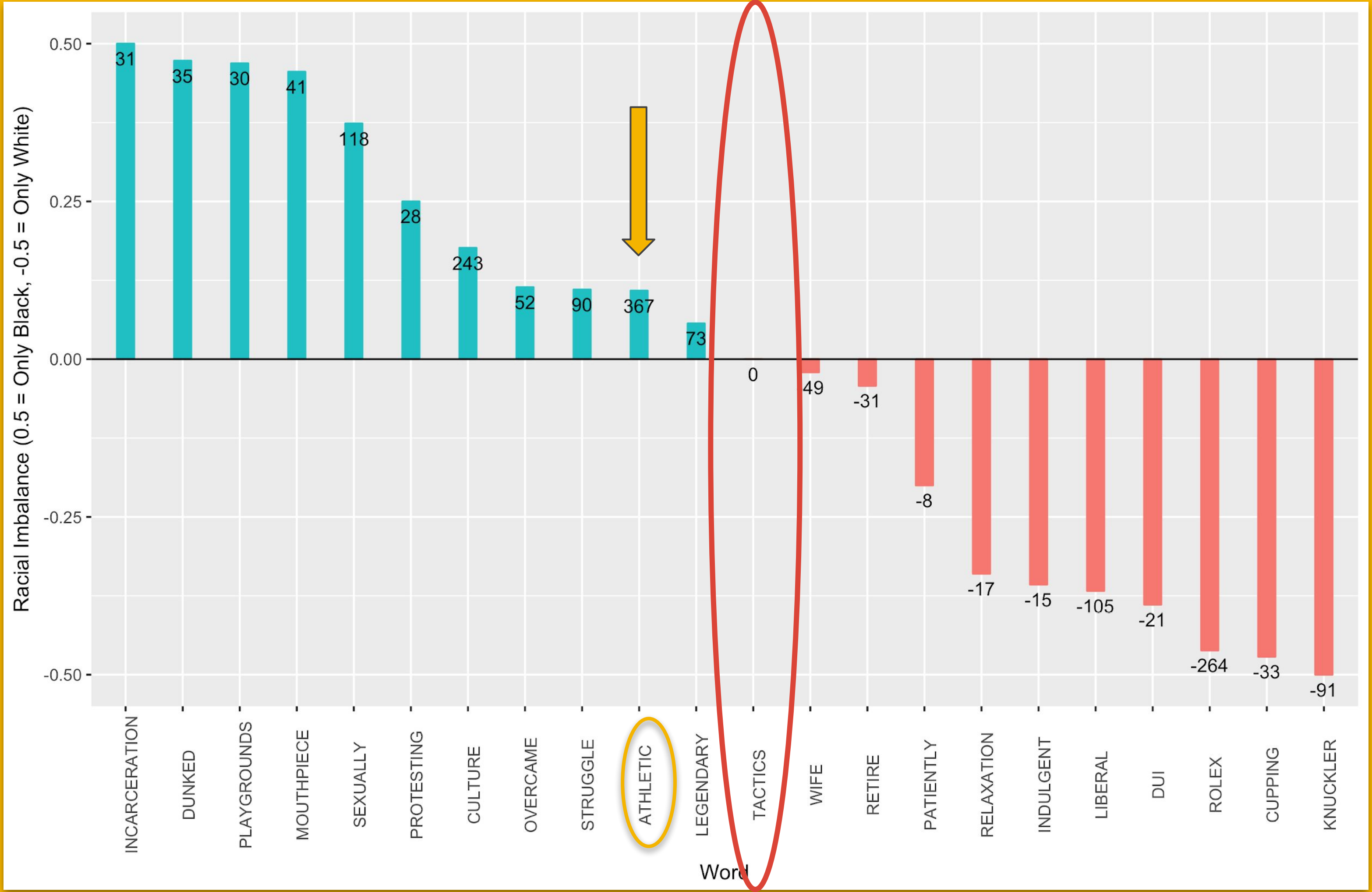


Tactics
43:43

Black Subcorpus →

Balanced across subcorpora

White Subcorpus →

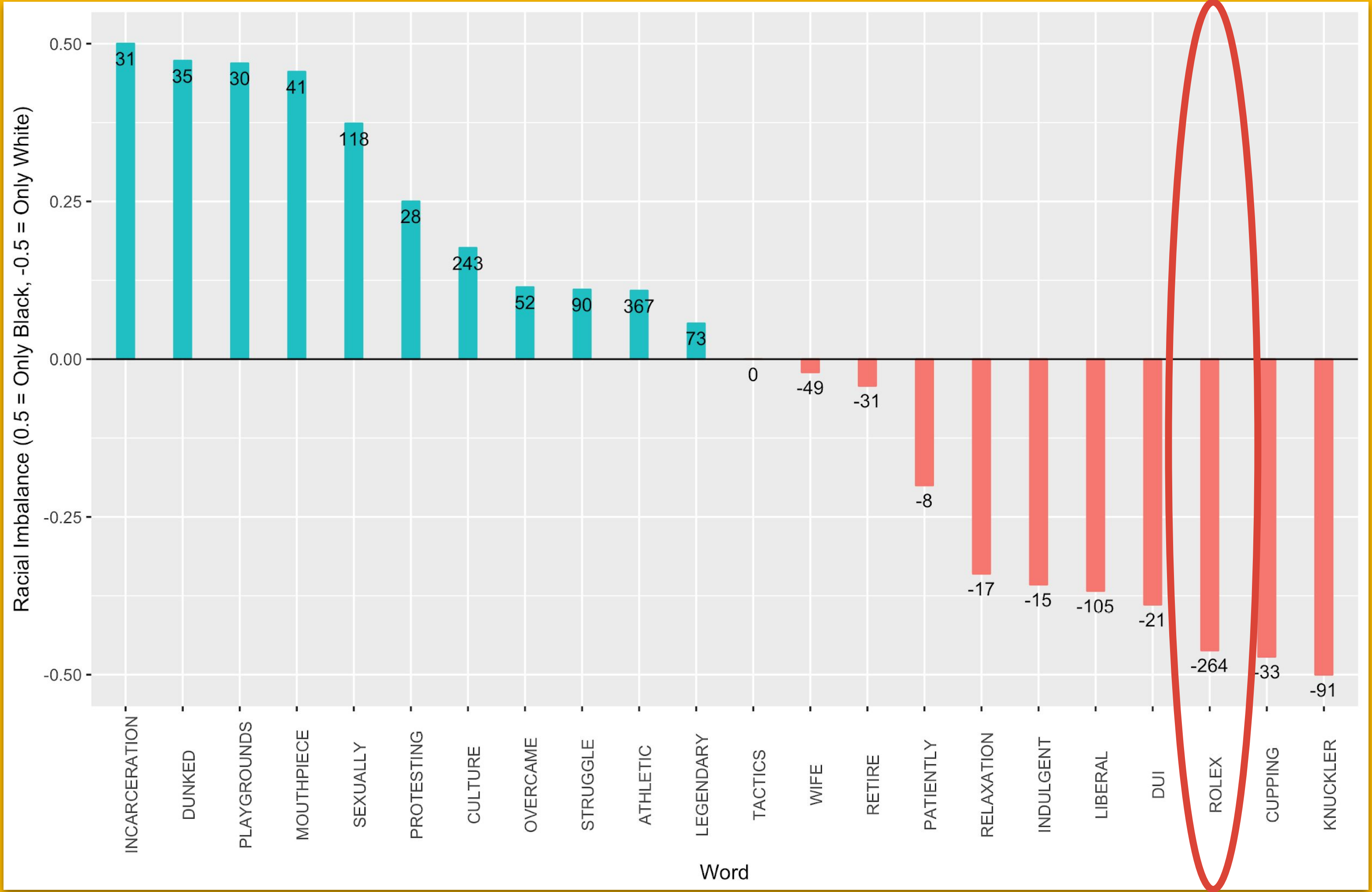


Rolex
11:275

Black Subcorpus →

96% in White Subcorpus

White Subcorpus →



Machine Learning

Step Two
Analyzing Counted Stuff

- A Random Forest is a learning algorithm building multiple decisions trees.
  - Modeling based on the most accurate.
  - Trained on the same dataset.

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 athletes by probability of membership into 2 groups
  - Based on sheer lexical frequency
  - Sorted by Race
- The most impressive result here is a lack of gradience in the probabilities of category membership.
  - The algorithm was sure. Like, super sure.

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.
  - Decision trees, tested on actual groupings.

Apparent Time

- Organize corpus into 20 year diachronic chunks.
- Repeat analysis and map semantic field
  - À la Caliskan et al. (2017)

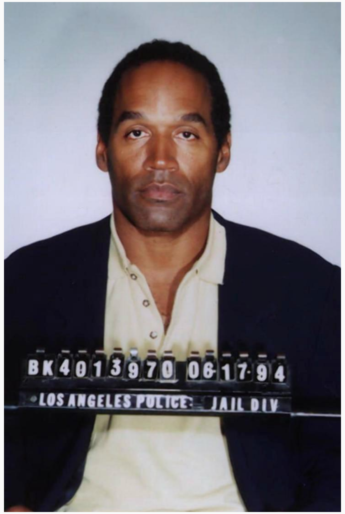
Apparent Time

- Watch as positions change over time
- Uncover the importance of social events or individual actors to racialization outcomes in this written context

Notoriety

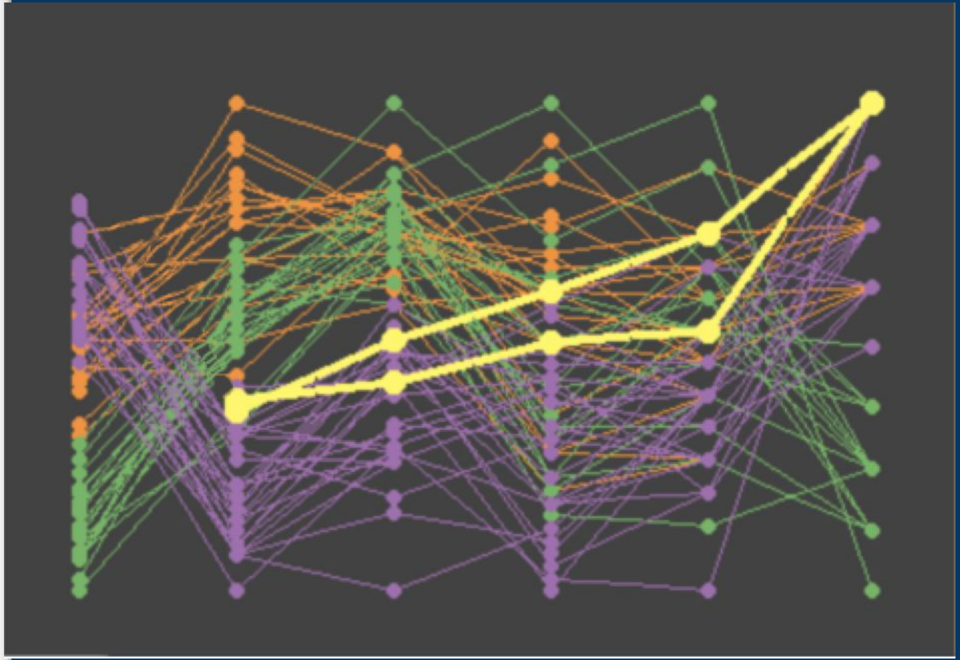
- Jackie Joyner Kersee
- Born 1962
- Olympic Gold Medalist 1988
- While racialized terms are used to describe her, **we don’t see** an increase in semantic intensity, as with Serena Williams

Events



- ❖ The Juice vs OJ Simpson
  - ➢ The greatest running back in living memory (1973)
  - ➢ Murderer? (1994)
- ❖ Semantic Intensity **ceiling** of racialized terms **increased**
  - ➢ *Savage*
  - ➢ *Insatiable*

Parallel Coordinates Plot



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**."

Caliskan et al. (2017) Semantics derived automatically from language corpora contain human-like biases

Thanks!

Kelly E. Wright

@raciolinguistic

kellywri@umich.edu

UNIVERSITY OF MICHIGAN