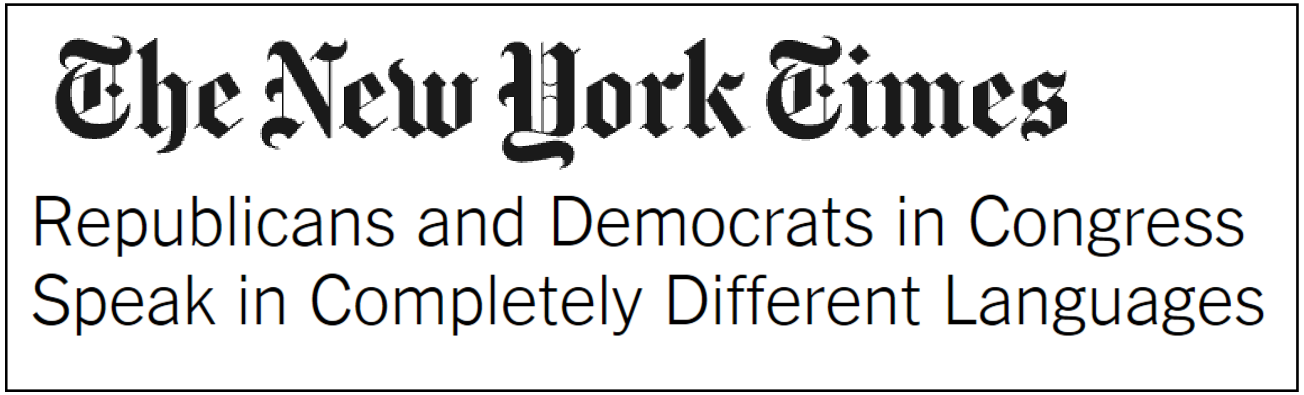
Machine Learning
+
Social Science

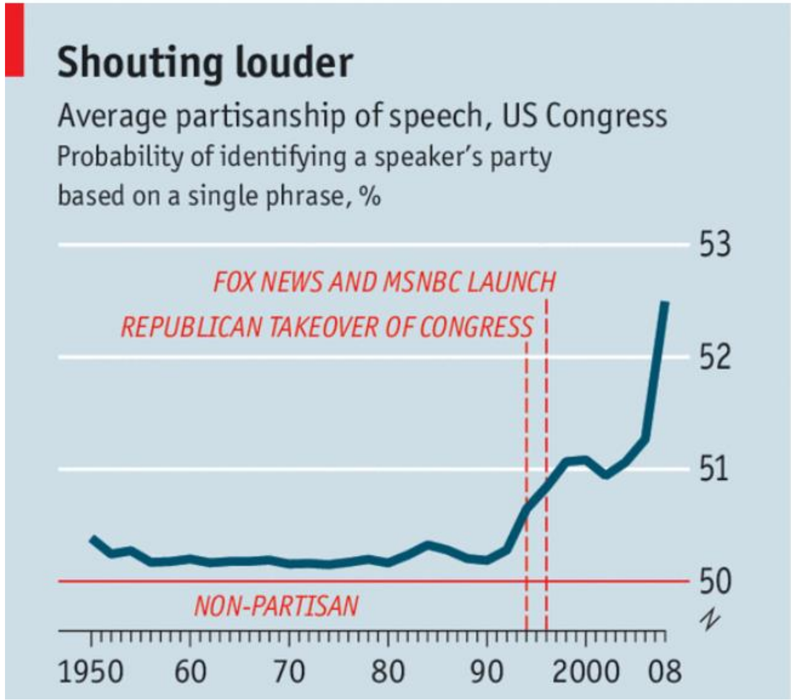
Matt Taddy — Microsoft Research and Chicago Booth



The New York Times

Republicans and Democrats in Congress Speak in Completely Different Languages

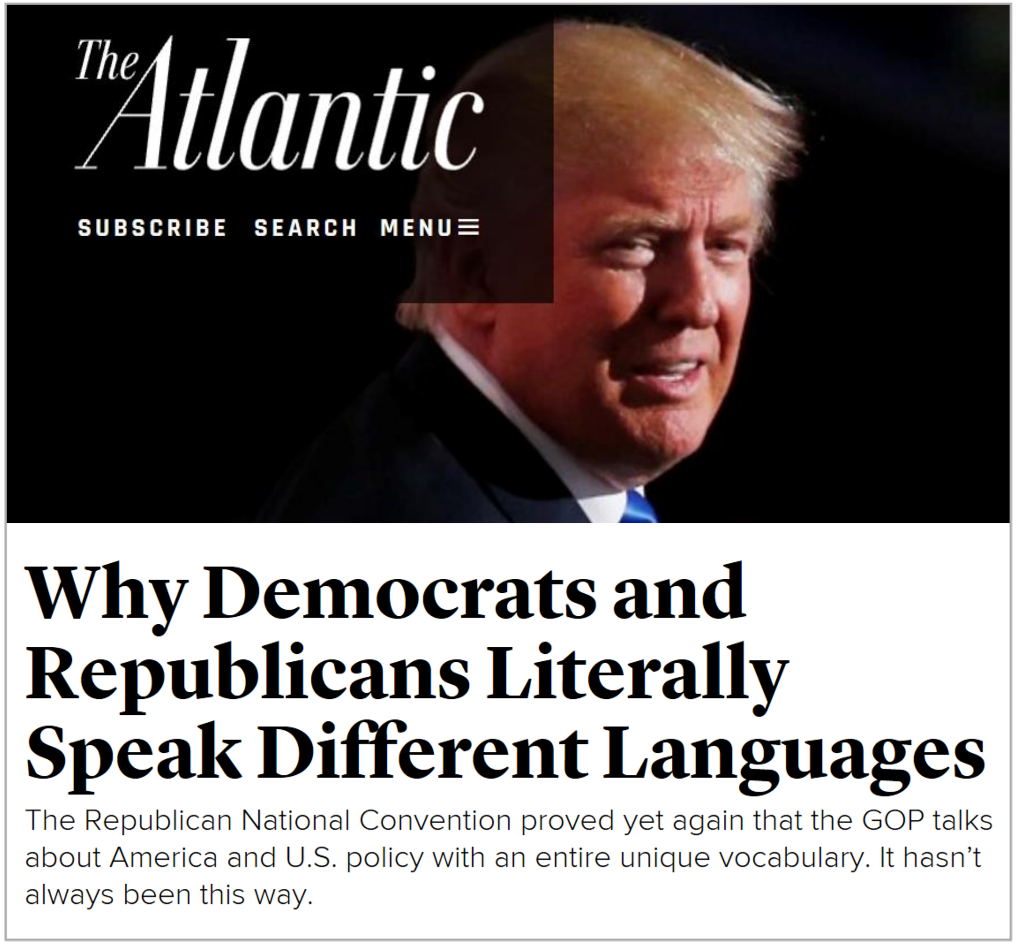
The Economist



Shouting louder

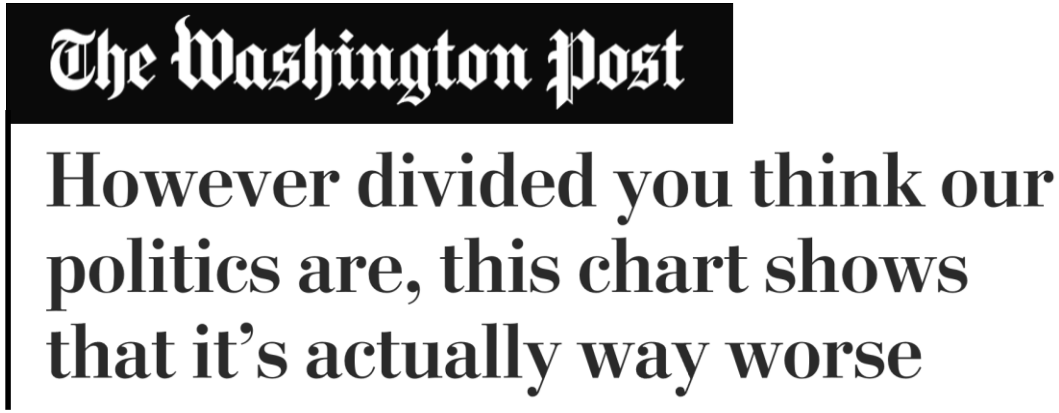
Average partisanship of speech, US Congress

Probability of identifying a speaker's party based on a single phrase, %



Why Democrats and Republicans Literally Speak Different Languages

The Republican National Convention proved yet again that the GOP talks about America and U.S. policy with an entire unique vocabulary. It hasn't always been this way.



The Washington Post

However divided you think our politics are, this chart shows that it's actually way worse

Wealthiest

Pro life

Tax Freedom

entrepreneurs

1 percent

Tax Relief

Tax Breaks

fair labor

Pro choice

ESTATE TAX

freedom fighters

War on Terror

terrorists

DEATH TAX

equality

Welfare Queens

Right to life

undocumented worker

illegal alien

Big Government

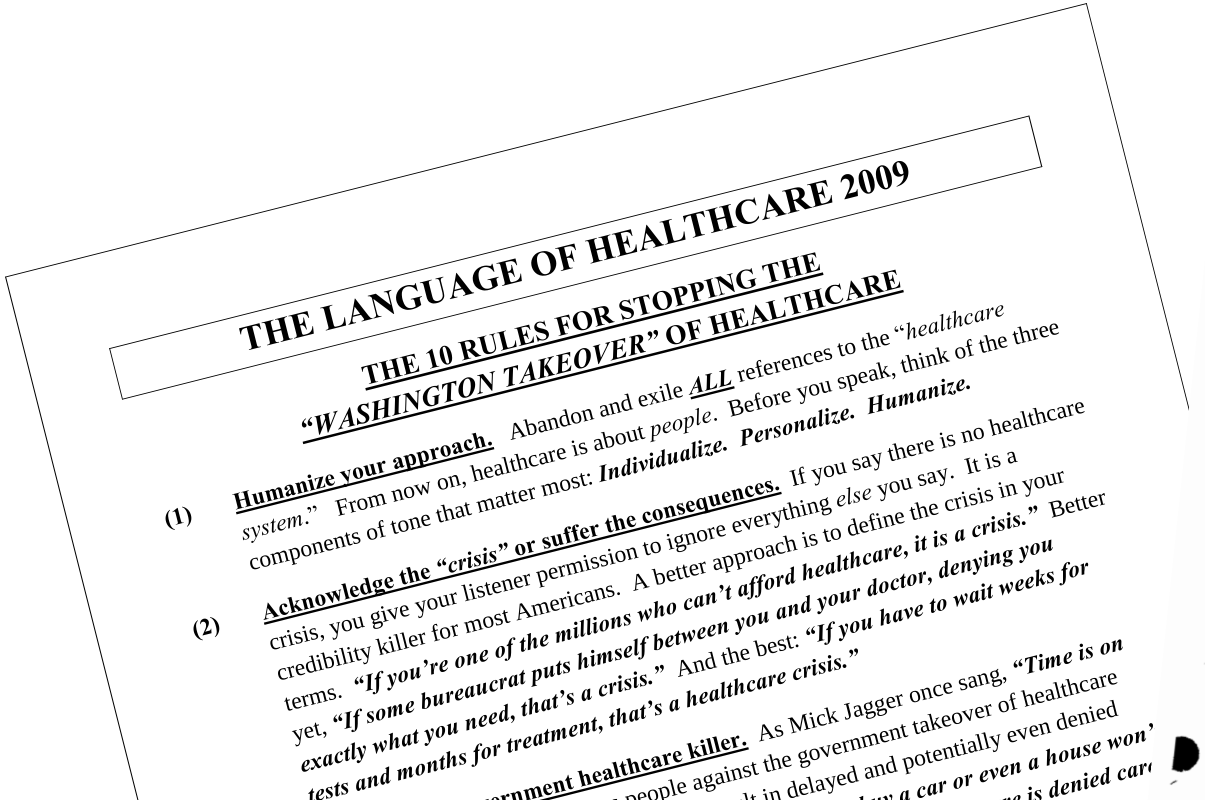
living wage

African American

capitalist

Washington takeover

Luntz (2006): “Never say '**privatization / private accounts**.' Instead say '**personalization / personal accounts**.' Two-thirds of America want to personalize security while only one third would privatize it. Why? [Personalization] suggests ownership and control... while [privatization] suggests a profit motive and winners and losers.”



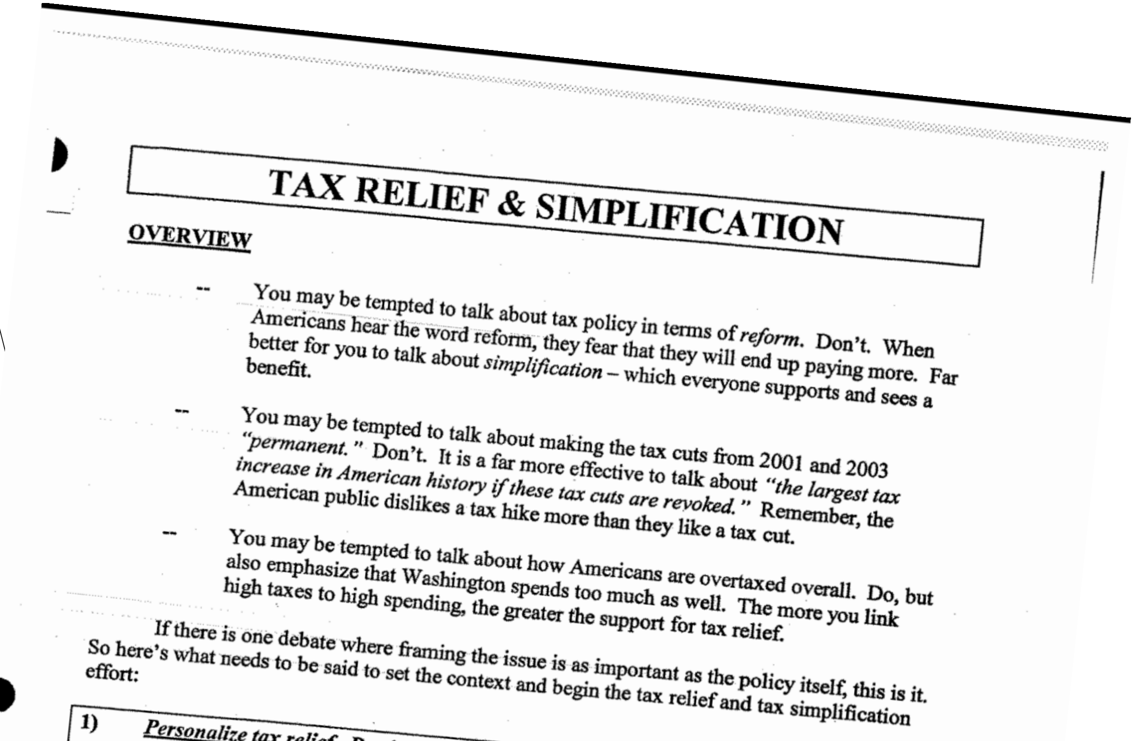
THE LANGUAGE OF HEALTHCARE 2009

<u>THE 10 RULES FOR STOPPING THE</u>
<u>“*WASHINGTON TAKEOVER*” OF HEALTHCARE</u>

(1) **<u>Humanize your approach.</u>** Abandon and exile ***ALL*** references to the “*healthcare system*.” From now on, healthcare is about *people*. Before you speak, think of the three components of tone that matter most: ***Individualize. Personalize. Humanize.***

(2) **<u>Acknowledge the “*crisis*” or suffer the consequences.</u>** If you say there is no healthcare crisis, you give your listener permission to ignore everything *else* you say. It is a credibility killer for most Americans. A better approach is to define the crisis in your terms. ***“If you’re one of the millions who can’t afford healthcare, it is a crisis.”*** Better yet, ***“If some bureaucrat puts himself between you and your doctor, denying you exactly what you need, that’s a crisis.”*** And the best: ***“If you have to wait weeks for tests and months for treatment, that’s a healthcare crisis.”***

<u>vernment healthcare killer.</u> As Mick Jagger once sang, ***“Time is on***
people against the government takeover of healthcare
lt in delayed and potentially even denied
uy a car or even a house won’
e is denied car



TAX RELIEF & SIMPLIFICATION

<u>OVERVIEW</u>

-- You may be tempted to talk about tax policy in terms of *reform*. Don’t. When Americans hear the word reform, they fear that they will end up paying more. Far better for you to talk about *simplification* – which everyone supports and sees a benefit.

-- You may be tempted to talk about making the tax cuts from 2001 and 2003 *“permanent.”* Don’t. It is a far more effective to talk about *“the largest tax increase in American history if these tax cuts are revoked.”* Remember, the American public dislikes a tax hike more than they like a tax cut.

-- You may be tempted to talk about how Americans are overtaxed overall. Do, but also emphasize that Washington spends too much as well. The more you link high taxes to high spending, the greater the support for tax relief.

If there is one debate where framing the issue is as important as the policy itself, this is it. So here’s what needs to be said to set the context and begin the tax relief and tax simplification effort:

1) ***<u>Personalize tax relief.</u>***

Data

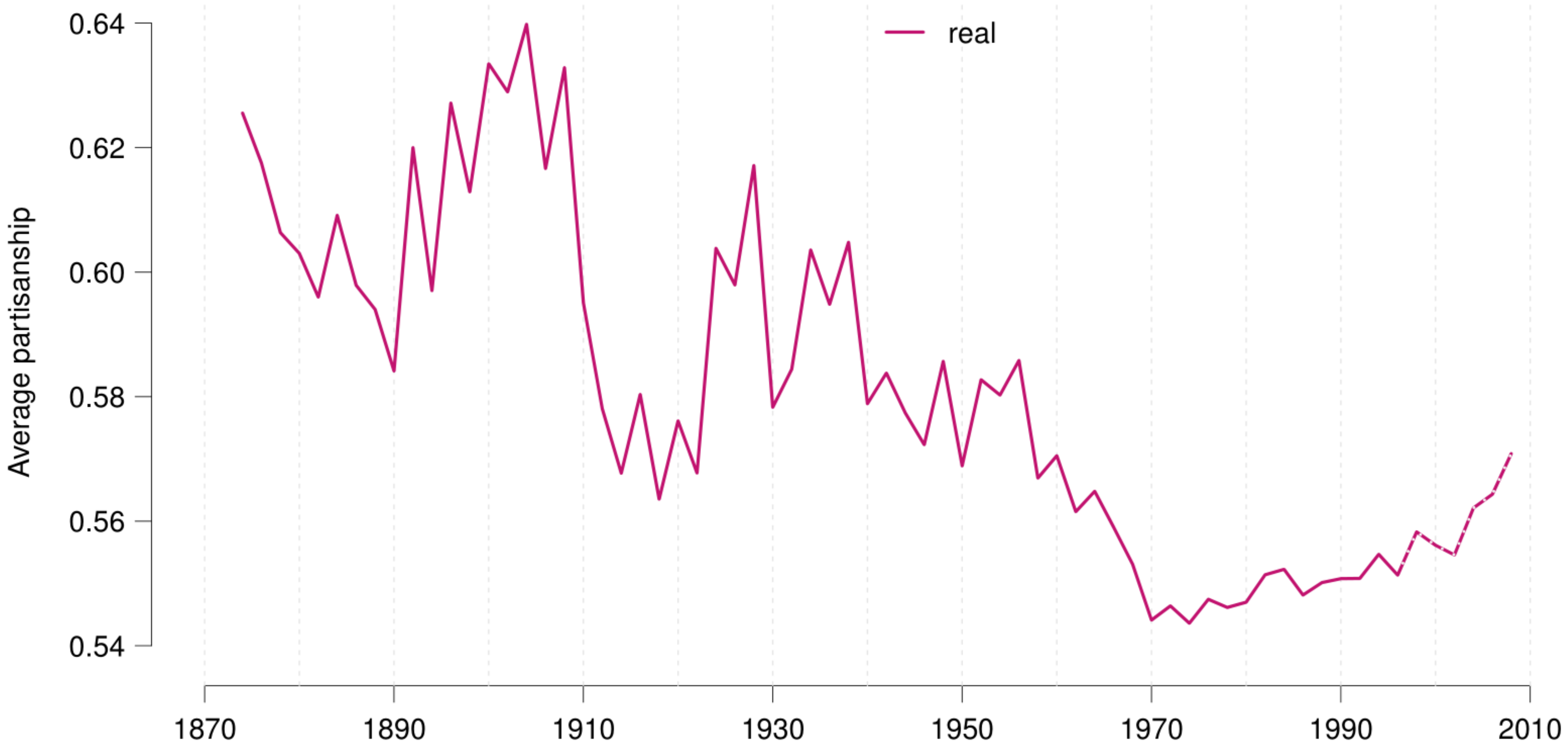
- US Congressional Record, 1873-2009
- Tokens (e.g. “war terror”) by speaker-session

$$\text{prob}(\textit{word } j \text{ by } \textit{speaker } i \text{ at } \textit{time } t) = q_{tj}(\boldsymbol{x}_{it}, P_i)$$

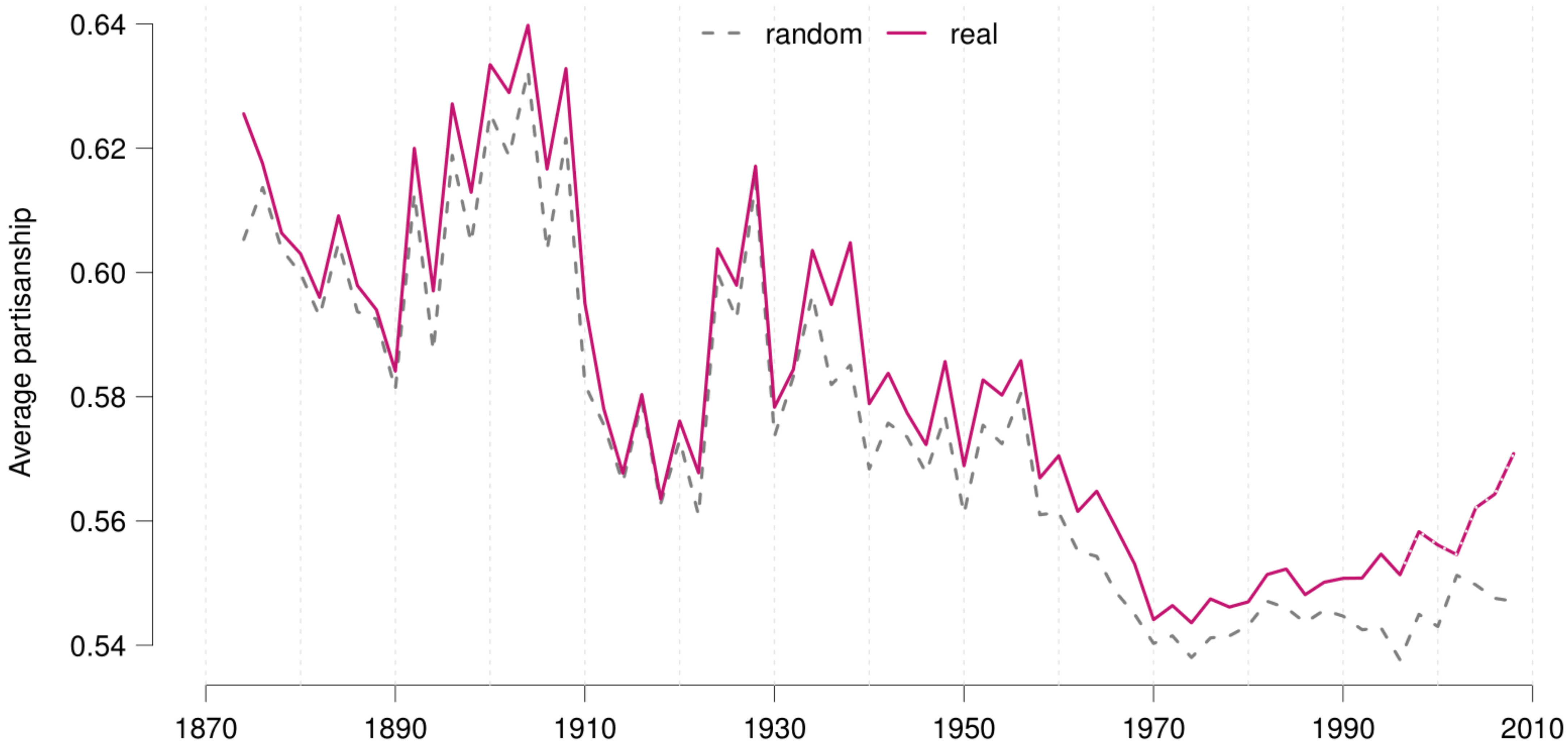
Model

- A strategic model for language choice given party P
- Token counts are multinomial, with probabilities

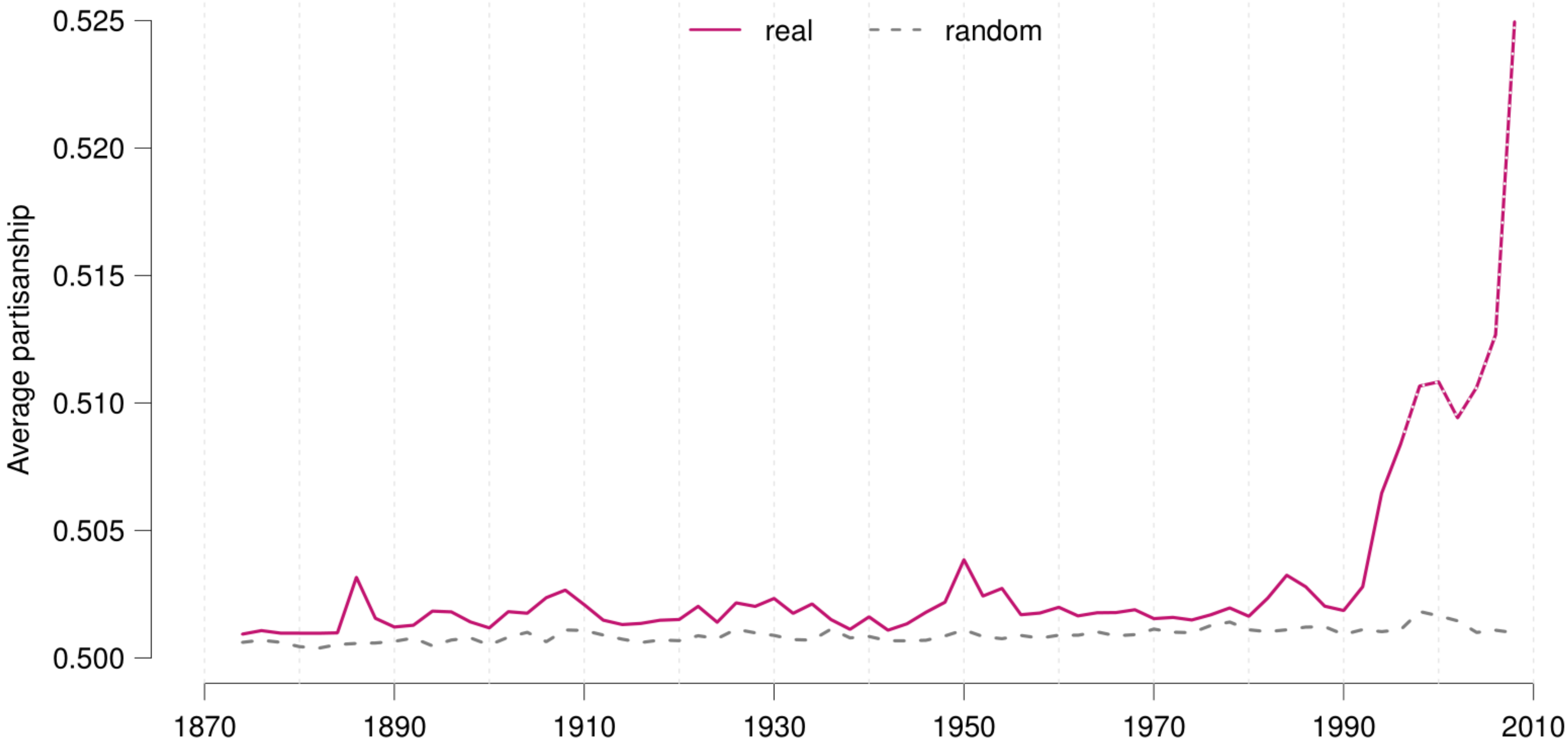
Naïve Estimation (MLE)



Naïve Estimation (MLE)



Machine Learning Estimator



Why is this hard?

Old tools don’t work in new data domains, but off-the-shelf ML is not directed towards the “Why?” questions of social science.

You need to take from everywhere

- Economics for a measurable and interpretable strategic model
- ML for reliable estimation with unstructured ultra-HD data
- Mathematical Statistics for uncertainty quantification

Artificial Intelligence and Economics

We’ve created NLP tools for questions on science, systems, + causation. Similar results in classic `Big Data’ environments (e.g., browser history)

We’re looking to replicate this success with Deep Neural Nets on all sorts of data, including richer language and images and even video.

THANKS!