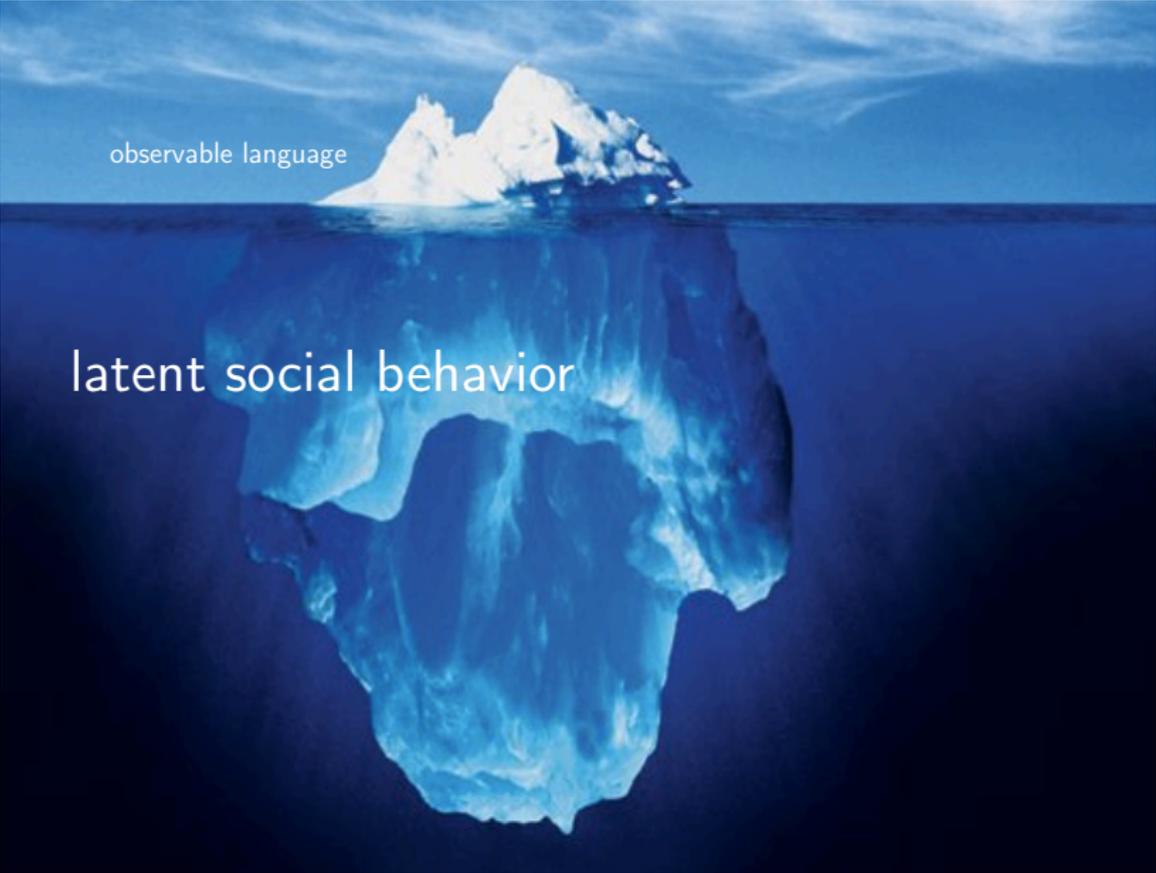


# Measuring Ideological Proportions in Political Speeches

Noah Smith

University of Washington

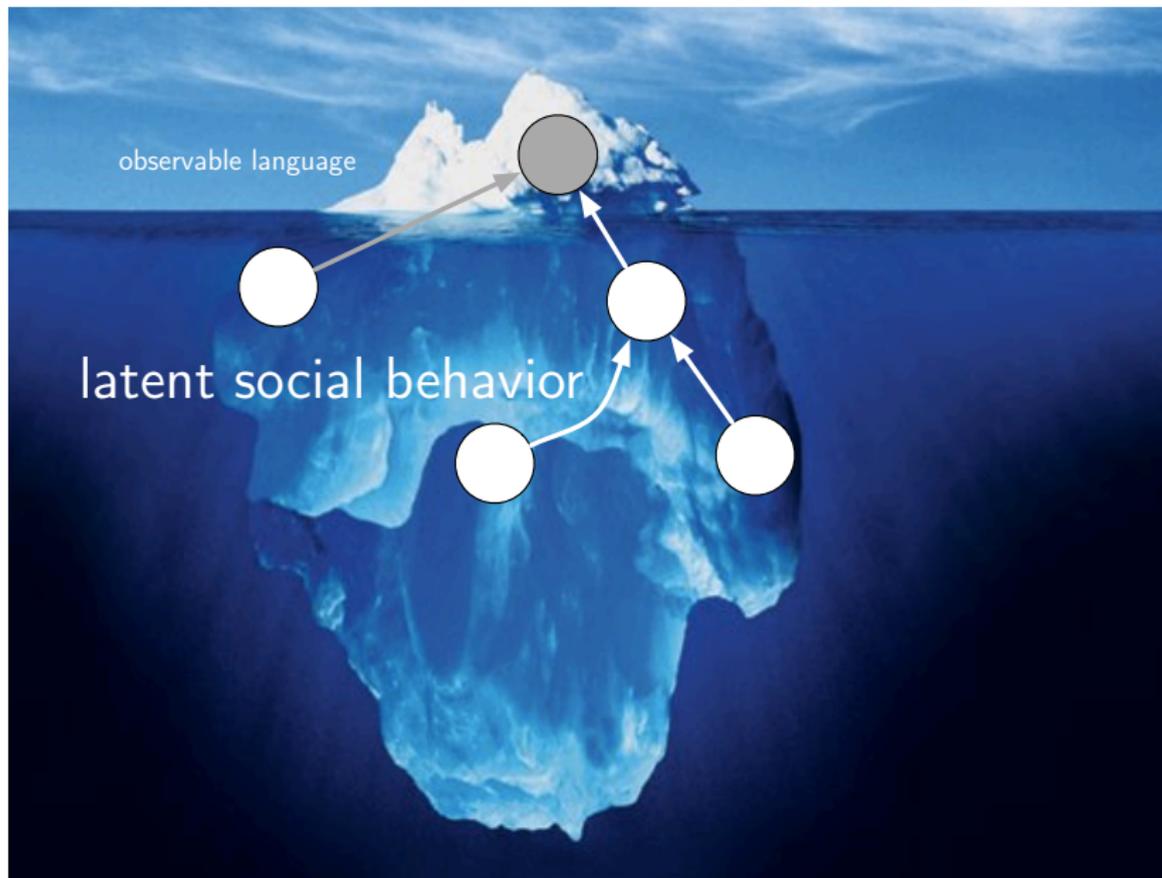
June 24, 2015



observable language

The image features a large iceberg floating in a deep blue ocean under a clear sky. The top of the iceberg, which is above the water line, is relatively small and jagged. The much larger, submerged portion of the iceberg is visible through the water, showing a complex, multi-faceted shape. The text 'observable language' is positioned above the water line, and 'latent social behavior' is positioned below it, illustrating the metaphor.

latent social behavior



# Measuring Ideological Proportions



“Well, I think you hit a reset button for the fall campaign. Everything changes. It’s almost like an Etch-A-Sketch. You can kind of shake it up and restart all over again.”

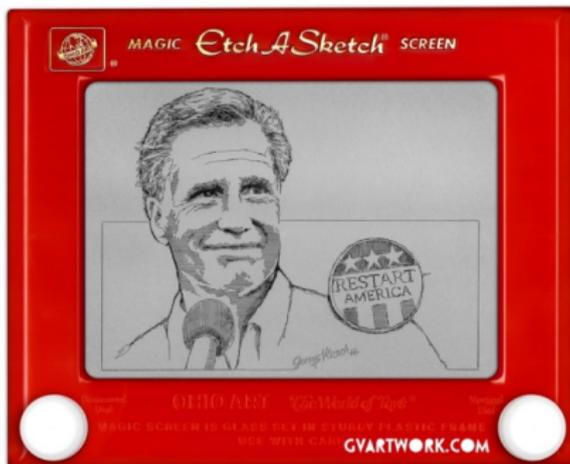
—Eric Fehrstrom, Mitt Romney’s spokesman, 2012

# Measuring Ideological Proportions



“Well, I think you hit a reset button for the fall campaign. Everything changes. It’s almost like an Etch-A-Sketch. You can kind of shake it up and restart all over again.”

—Eric Fehrstrom, Mitt Romney’s spokesman, 2012



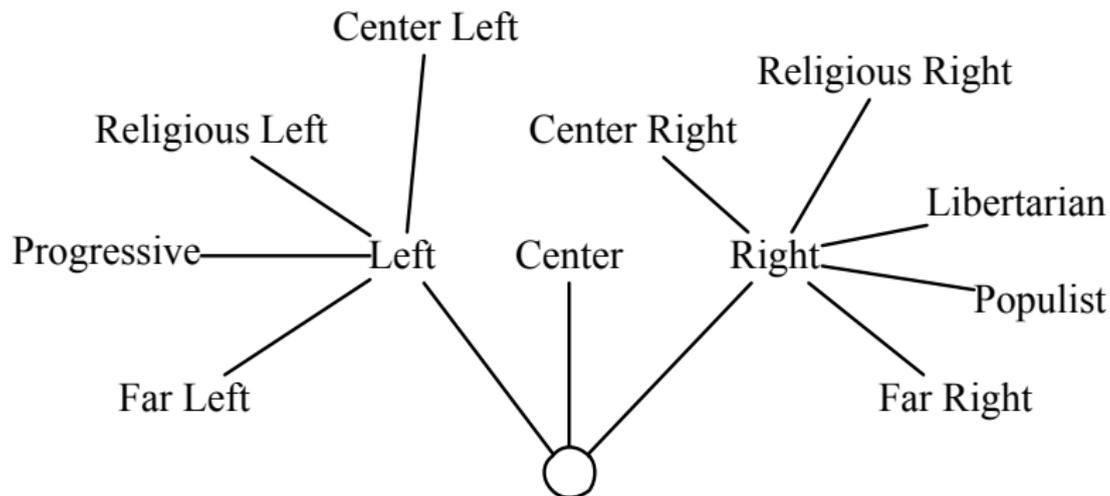
# The “Etch-a-Sketch” Hypothesis



Presidential candidates (rhetorically, at least) reposition themselves to be less extreme after winning a primary election.

- ▶ Predicted by the median voter theorem (Hotelling, 1929; Black, 1948; Downs, 1957)

# Ideology Typology





Instead of putting more limits on your earnings and your options, we need to place clear and firm limits on **government spending**. As a start, I will lower **federal spending** to 20 percent of GDP within four years' time – down from the 24.3 percent today.

The President's plan assumes an endless expansion of government, with costs rising and rising with the spread of Obamacare. I will halt the expansion of government, and **repeal Obamacare**.

Working together, we can save **Social Security** without making any changes in the system for people in or nearing retirement. We have two basic options for future retirees: a **tax increase** for high-income retirees, or a decrease in the benefit **growth rate** for high-income retirees. I favor the second option; it protects everyone in the system and it avoids **higher taxes** that will drag down the economy

I have proposed a Medicare plan that improves the program, keeps it solvent, and slows the rate of growth in **health care costs**.

—Excerpt from speech by Romney on 5/25/12 in Des Moines, IA



Instead of putting more limits on your earnings and your options, we need to place clear and firm limits on **government spending**. As a start, I will lower **federal spending** to 20 percent of GDP within four years' time – down from the 24.3 percent today.

The President's plan assumes an endless expansion of government, with costs rising and rising with the spread of Obamacare. I will halt the expansion of government, and **repeal Obamacare**.

Working together, we can save **Social Security** without making any changes in the system for people in or nearing retirement. We have two basic options for future retirees: a **tax increase** for high-income retirees, or a decrease in the benefit **growth rate** for high-income retirees. I favor the second option; it protects everyone in the system and it avoids **higher taxes** that will drag down the economy.

I have proposed a Medicare plan that improves the program, keeps it solvent, and slows the rate of growth in **health care costs**.

—Excerpt from speech by Romney on 5/25/12 in Des Moines, IA

# Cue-Lag Representation of a Speech



20 government spending 8 federal spending 47 repeal Obamacare 7

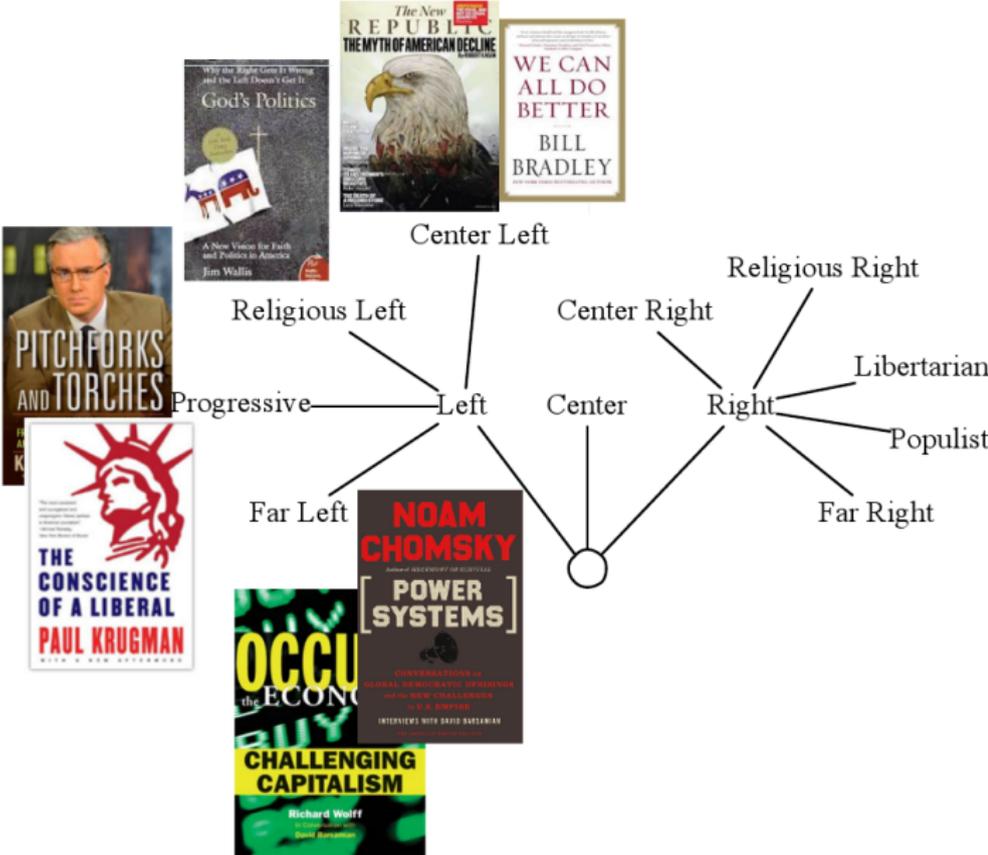
Social Security 24 tax increase 13 growth rate 21 higher taxes 29

health care costs

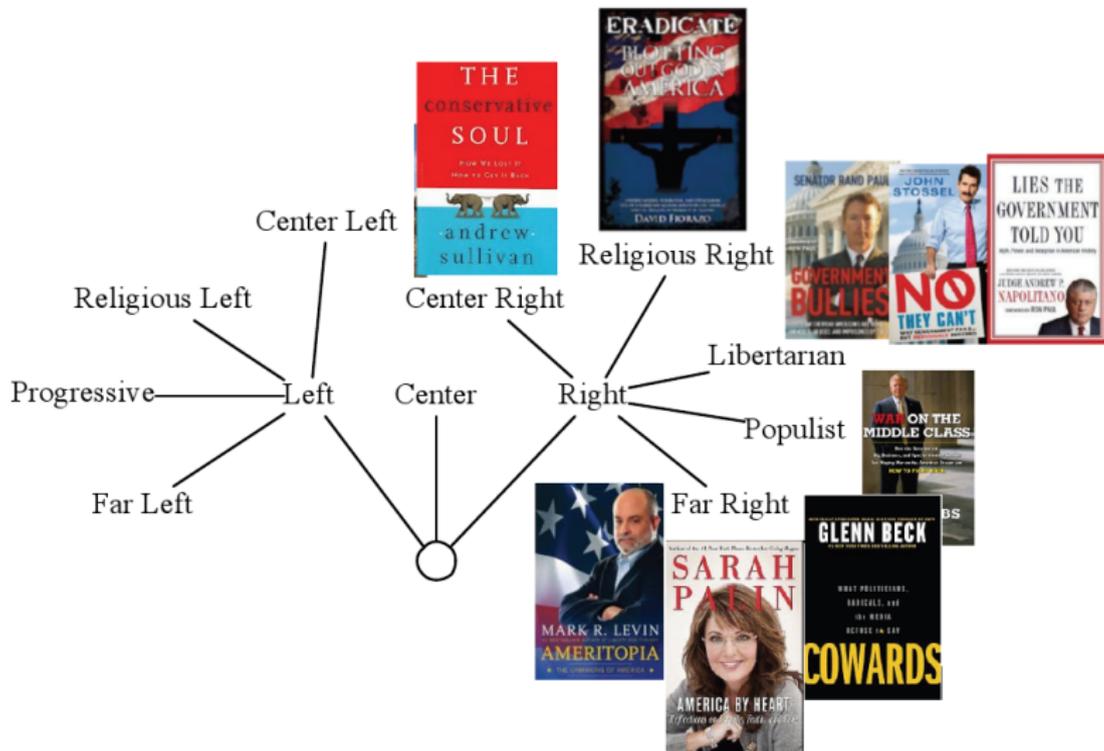


1. Infer a lexicon of “cue terms” using text associated with known ideologies.
2. Probabilistic reasoning about the ideological intent of each cue term in context.

# Ideological Books Corpus



# Ideological Books Corpus



# Step 1: Cue Lexicon



For every term  $t$  (i.e., a bigram or trigram):

$$\log p \left( t \mid \text{[Book Cover: NOAM CHOMSKY POWER SYSTEMS]} \right) \propto \begin{matrix} \eta_t^{\text{background}} \\ + \eta_t^{\text{Left}} \\ + \eta_t^{\text{Far Left}} \\ + \eta_t^{\text{education topic}} \\ + \eta_t^{\text{Power Systems}} \end{matrix}$$

- ▶ A sparse prior is placed on each effect  $\eta_t^*$  (Eisenstein et al., 2011).
- ▶ After estimation, a term  $t$  is associated with “Far Left” if  $\eta_t^{\text{Far Left}} > 0$ .

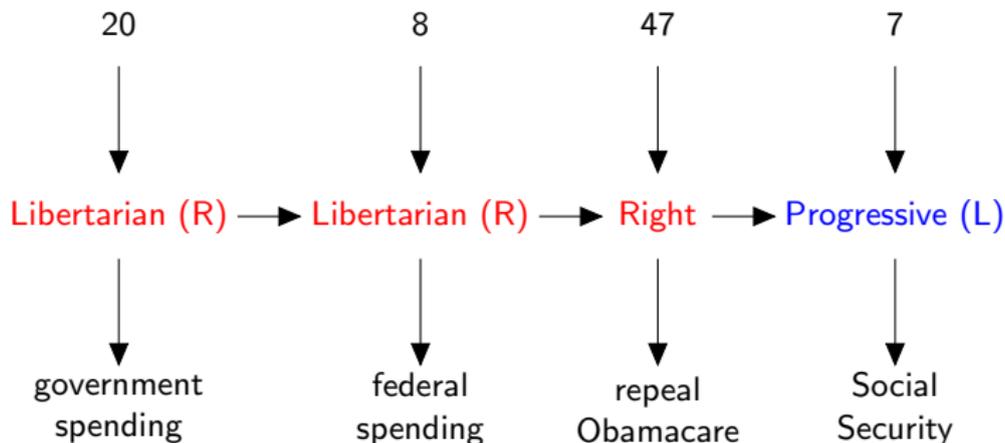
# Example Cues



<b>Center-Right</b> D. Frum, M. McCain, C. T. Whitman (1,450)	governor bush; class voter; health care; republican president; george bush; state police; move forward; miss america; middle eastern; water buffalo; fellow citizens; sam's club; american life; working class; general election; culture war; status quo; human dignity; same-sex marriage
<b>Libertarian</b> Rand Paul, John Stossel, <i>Reason</i> (2,268)	medical marijuana; raw milk; rand paul; economic freedom; health care; government intervention; market economies; commerce clause; military spending; government agency; due process; drug war; minimum wage; federal law; ron paul; private property
<b>Religious Right</b> (960)	daily saint; holy spirit; matthew [c/v]; john [c/v]; jim wallis; modern liberals; individual liberty; god's word; jesus christ; elementary school; natural law; limited government; emerging church; private property; planned parenthood; christian nation; christian faith

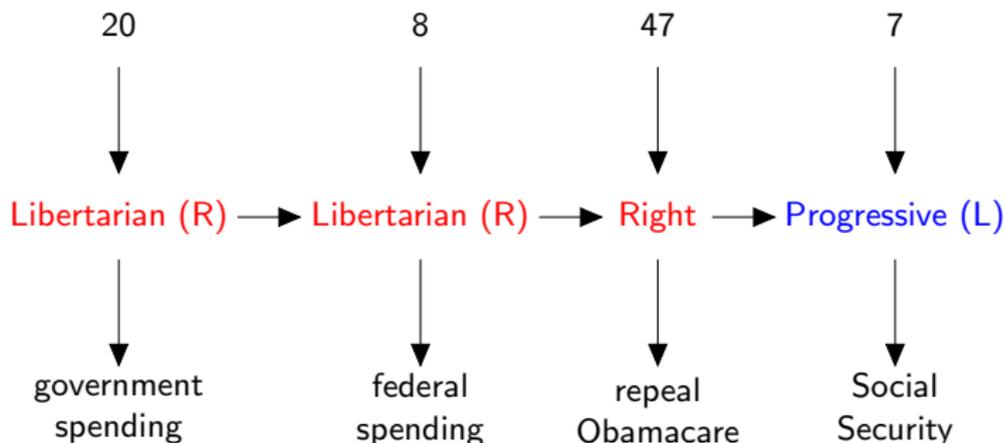
Browse results at <http://www.ark.cs.cmu.edu/CLIP/>.

## Step 2: Cue-Lag Ideological Proportions Model



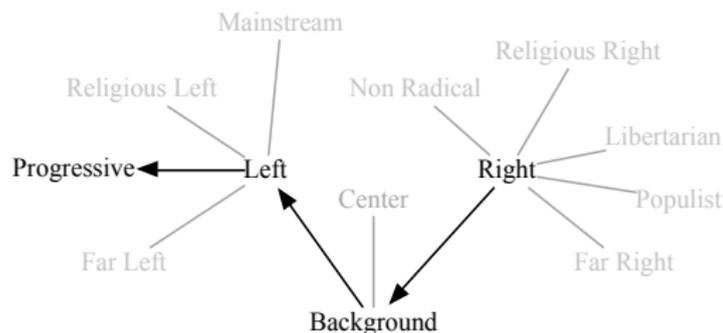
$$p(\text{Prog.} \mid \text{Right}, 7) = + \quad p\left(\text{RESET} \mid 7\right) \times p(\text{Back.} \rightsquigarrow \text{Prog.}) \\
 + \quad (1 - p\left(\text{RESET} \mid 7\right)) \times p(\text{Right} \rightsquigarrow \text{Prog.})$$

## Step 2: Cue-Lag Ideological Proportions Model



$$p(\text{Prog.} \mid \text{Right}, 7) = \frac{(1 - (1 - \rho)^{7+1}) \times p(\text{Back.} \rightsquigarrow \text{Prog.})}{(1 - \rho)^{7+1} \times p(\text{Right} \rightsquigarrow \text{Prog.})}$$

# Ideological “Transition”



$$p(\text{Right} \rightsquigarrow \text{Prog.}) = \begin{matrix} \times & (1 - \zeta_{\text{Right}}) & \times & \theta_{\text{Right} \rightarrow \text{Back.}} \\ \times & (1 - \zeta_{\text{Back.}}) & \times & \theta_{\text{Back.} \rightarrow \text{Left}} \\ \times & (1 - \zeta_{\text{Left}}) & \times & \theta_{\text{Left} \rightarrow \text{Prog.}} \\ \times & \zeta_{\text{Prog.}} & & \end{matrix}$$

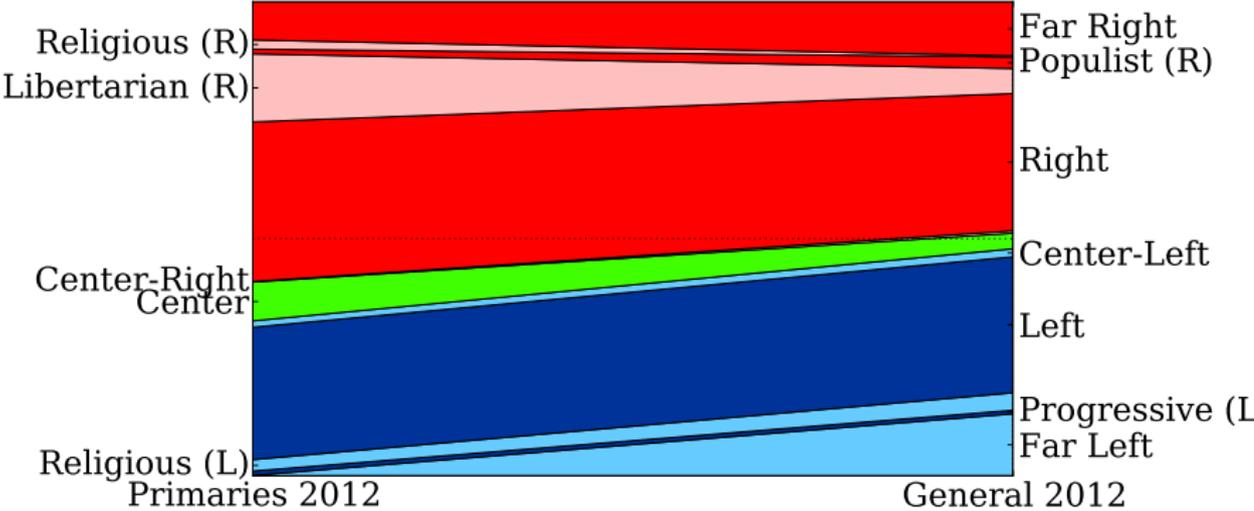


- ▶ Emission probabilities  $p(\text{term} \mid \text{ideology})$  are latent, drawn from an *asymmetric* Dirichlet.
- ▶ Collapsed Gibbs for ideology and restart variables
- ▶ Quantity of interest is ideological proportions:

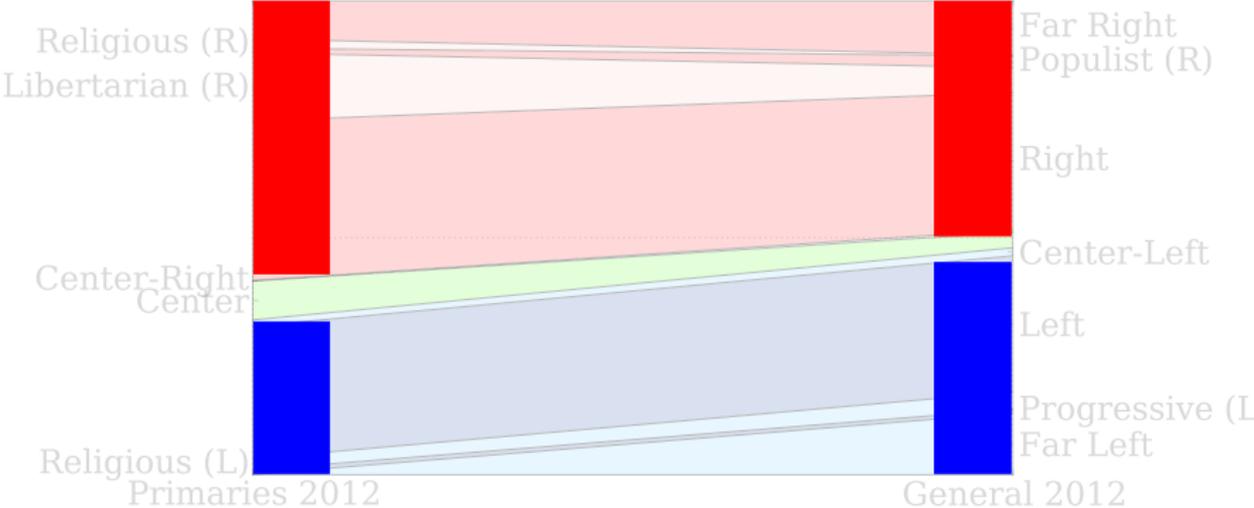
$$\sum_{t=1}^T p(\text{ideology at } t \mid \text{evidence, prior}) \frac{\frac{1}{2} (\text{lag before } t + \text{lag after } t)}{\text{length of speech}}$$

- ▶ 21 candidates, each separated into primary/general election epochs
  - ▶ Each candidate has his/her own emission distributions and separate transition distributions for each epoch

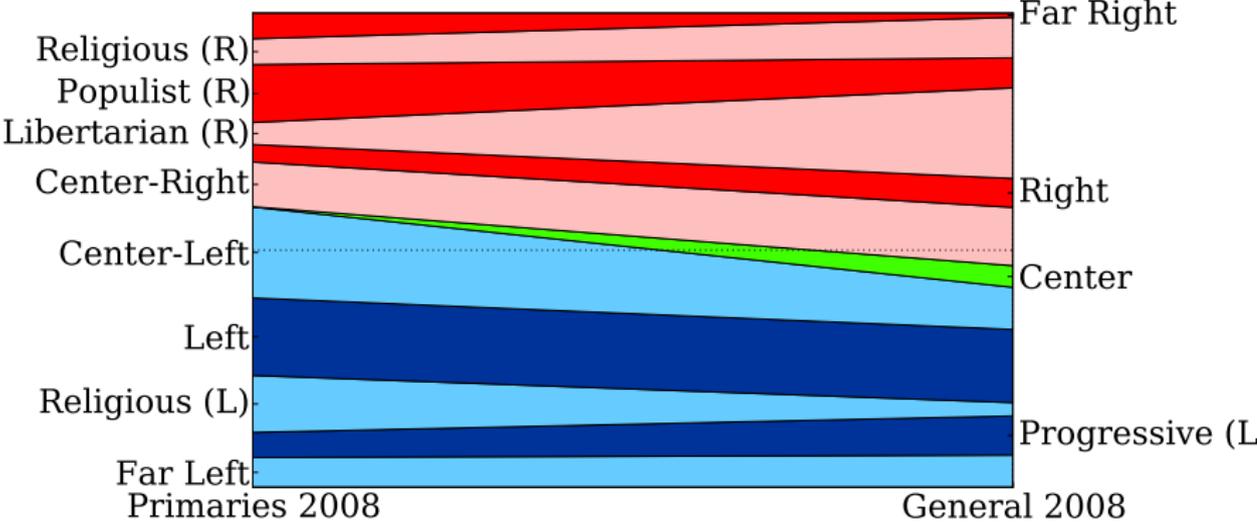
# Mitt Romney



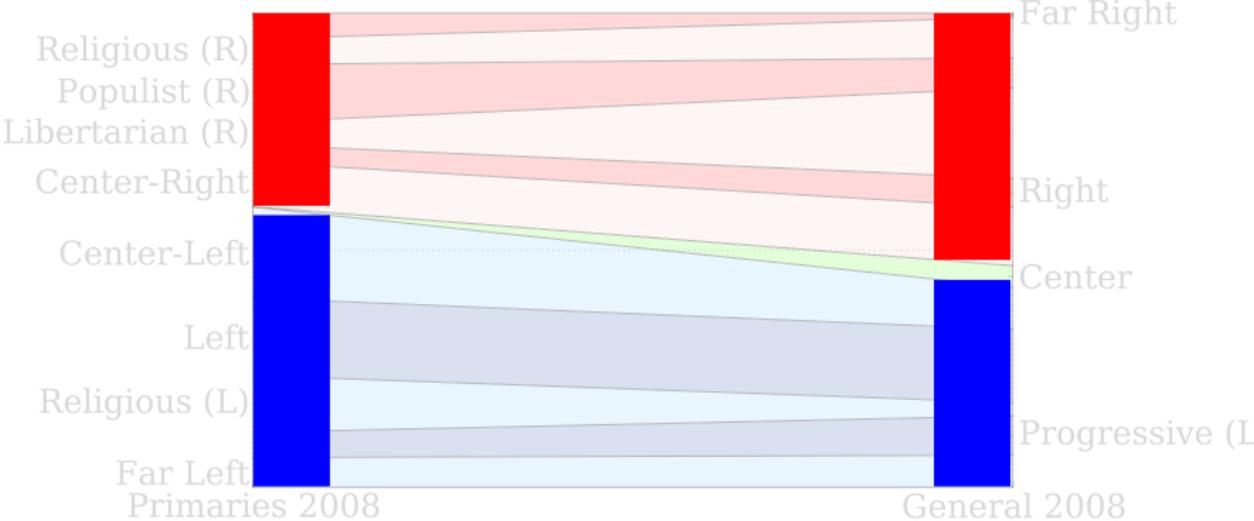
# Mitt Romney

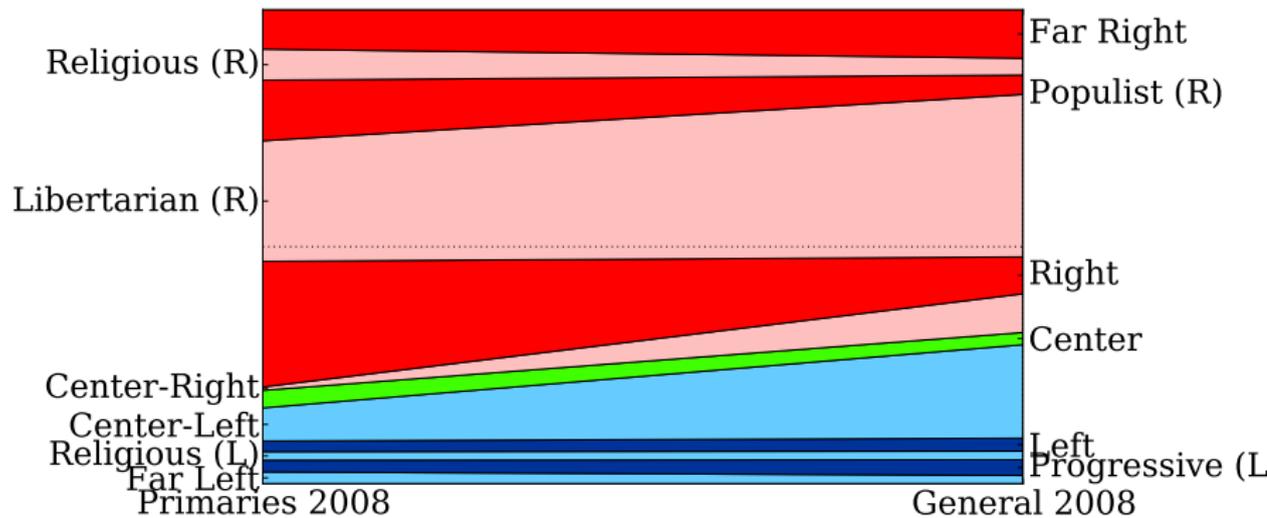


# Barack Obama

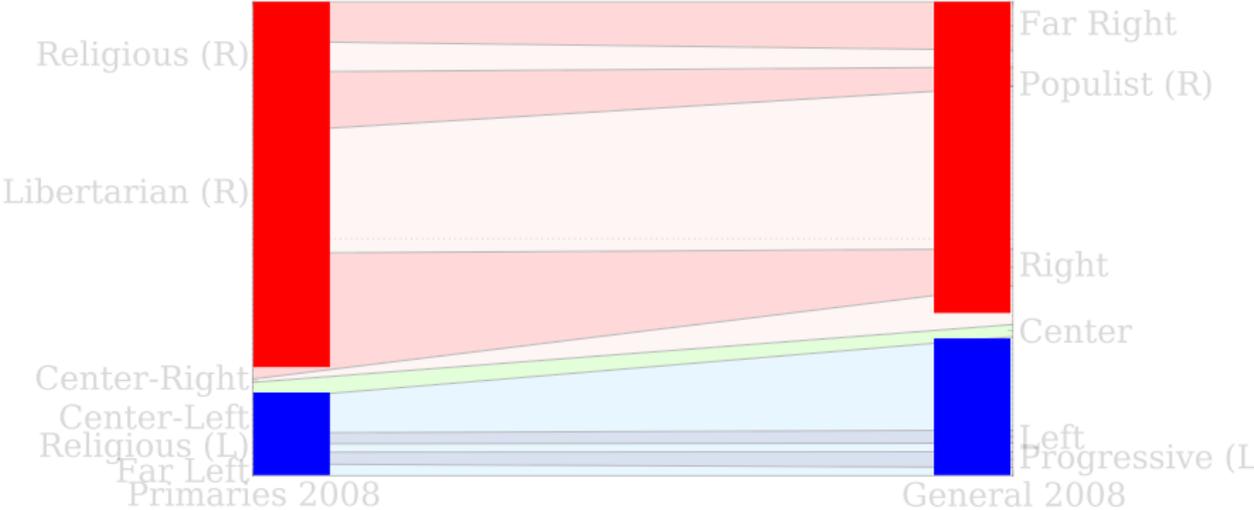


# Barack Obama





# John McCain



# Objective Evaluation?



## Pre-registered hypothesis

A statement by a domain expert about his/her *expectations* of the model's output.



## Hypotheses

### *Sanity checks (strong):*

- S1. Republican primary candidates should tend to draw more from RIGHT than from LEFT.
- S2. Democratic primary candidates should tend to draw more from LEFT than from RIGHT.
- S3. In general elections, Democrats should draw more from the LEFT than the Republicans and vice versa for the RIGHT.

### *Primary hypotheses (strong):*

- P1. Romney, McCain and other Republicans should almost never draw from FAR LEFT, and extremely rarely from PROGRESSIVE.
- P2. Romney should draw more heavily from the RIGHT than Obama in both stages of the 2012 campaign.

### *Primary hypotheses (moderate):*

- P3. Romney should draw more heavily on words from the LIBERTARIAN, POPULIST, RELIGIOUS RIGHT, and FAR RIGHT in the primary compared to the general election. In the general election, Romney should draw more heavily on CENTER, CENTER-RIGHT and LEFT vocabularies.



Compare against “simplified” versions of the model:

- ▶ HMM: traditional HMM without ideological tree structure
- ▶ NORES: weaker assumptions (never restart)
- ▶ MIX: stronger assumptions (always restart)

	CLIP	HMM	MIX	NORES
Sanity checks	<b>20/21</b>	19/22	21/22	17/22
Strong hypotheses	<b>31/34</b>	23/33	28/34	30/34
Moderate hypotheses	<b>14/17</b>	<b>14/17</b>	12/17	11/17
Total	<b>65/72</b>	56/72	61/73	58/73



- ▶ Predicting *responses* to language: was the speaker/author successful?
- ▶ Modeling *goals*: what can we infer about what an author wants?
- ▶ Offering strategic *advice*: what's the best way to frame a message?



Thank you!

Collaborators: Yanchuan Sim (CMU), Justin Gross, Brice Acree  
(UNC Chapel Hill)

More details: Sim et al. (2013)



- Black, D. (1948). On the rationale of group decision-making. *The Journal of Political Economy*, 56(1):23–34.
- Downs, A. (1957). *An Economic Theory of Democracy*. Harper, New York.
- Eisenstein, J., Ahmed, A., and Xing, E. P. (2011). Sparse additive generative models of text. In *Proceedings of ICML*.
- Hotelling, H. (1929). Stability in competition. *The Economic Journal*, 39(153):41–57.
- Sim, Y., Acree, B. D. L., Gross, J. H., and Smith, N. A. (2013). Measuring ideological proportions in political speeches. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Seattle, WA.