Measuring Ideological Proportions in Political Speeches

Noah Smith

University of Washington

June 24, 2015
observable language

latent social behavior
observable language

latent social behavior
“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 Fehrnstrom, 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 Fehrnstrom, 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

- Left
  - Religious Left
  - Progressive
  - Far Left

- Center
  - Center Left
  - Center Right

- Right
  - Religious Right
  - Libertarian
  - Populist
  - Far Right
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
Procedure

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

Center Left

Religious Left

Progressive

Far Left

Center

Right

Religious Right

Libertarian

Populist

Far Right
Ideological Books Corpus

- Center Left
- Religious Left
- Progressive
- Far Left
- Center
- Religious Right
- Libertarian
- Populist
- Far Right
- Right

Books:
- The Conservative Soul
- Eradicate Blocking Our Gods
- America
- Government Bullies
- Lies the Government Told You
- Lies They Can't
- A Man on the Middle Class
- Glenn Beck's Cowards
- Sarah Palin's Ameritopia
Step 1: Cue Lexicon

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

$$\log p(t) \propto \eta_{background} + \eta_{Left} + \eta_{Far Left} + \eta_{education topic} + \eta_{Power Systems}$$

- 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 $\eta^*_{Far Left} > 0$. 
## Example Cues

| **Center-Right** 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 |
| **Libertarian** Rand Paul, John Stossel, *Reason* (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 |
| **Religious Right** (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 |

Step 2: Cue-Lag Ideological Proportions Model

\[ p(\text{Prog.} | \text{Right}, 7) = + \]

\[ p(\text{Prog.} | 7) \times p(\text{Back.} \rightsquigarrow \text{Prog.}) \]

\[ (1 - p(\text{Prog.} | 7)) \times p(\text{Right} \rightsquigarrow \text{Prog.}) \]
Step 2: Cue-Lag Ideological Proportions Model

\[
p(\text{Prog.} \mid \text{Right, 7}) = (1 - (1 - \rho)^{7+1}) \times p(\text{Back.} \leadsto \text{Prog.}) + (1 - \rho)^{7+1} \times p(\text{Right} \leadsto \text{Prog.})
\]
Ideological “Transition”

\[ p(\text{Right} \rightsquigarrow \text{Prog.}) = (1 - \zeta_{\text{Right}}) \times (1 - \zeta_{\text{Back.}}) \times (1 - \zeta_{\text{Left}}) \times \zeta_{\text{Prog.}} \times \theta_{\text{Right} \rightarrow \text{Back.}} \times \theta_{\text{Back.} \rightarrow \text{Left}} \times \theta_{\text{Left} \rightarrow \text{Prog.}} \]
Inference

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

Religious (R)
Libertarian (R)
Center-Right
Center
Religious (L)
Primaries 2012
General 2012
Far Right
Populist (R)
Right
Center-Left
Left
Progressive (L)
Far Left
Barack Obama

Primaries 2008
- Far Left
- Religious (L)
- Left
- Center-Left
- Center-Right
- Libertarian (R)
- Populist (R)
- Religious (R)

General 2008
- Progressive (L)
- Center
- Right
- Far Right
Barack Obama

Primaries 2008
Far Left
Religious (L)
Left
Center-Left
Center-Right
Libertarian (R)
Populist (R)
Religious (R)
Progressive (L)
Center
Right
Far Right

General 2008
John McCain

Primaries 2008

General 2008

Far Left
Religious (L)
Center-Left
Center-Right
Libertarian (R)
Religious (R)
Progressive (L)

Left
Center
Right
Populist (R)

Far Right
Populist (R)
Right
Center
Left
Progressive (L)

Primaries 2008

General 2008
Objective Evaluation?

Pre-registered hypothesis
A statement by a domain expert about his/her expectations of the model’s output.
## Preregistered Hypotheses

<table>
<thead>
<tr>
<th>Hypotheses</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sanity checks (strong):</strong></td>
</tr>
<tr>
<td>S1. Republican primary candidates should tend to draw more from <strong>RIGHT</strong> than from <strong>LEFT</strong>.</td>
</tr>
<tr>
<td>S2. Democratic primary candidates should tend to draw more from <strong>LEFT</strong> than from <strong>RIGHT</strong>.</td>
</tr>
<tr>
<td>S3. In general elections, Democrats should draw more from the <strong>LEFT</strong> than the Republicans and vice versa for the <strong>RIGHT</strong>.</td>
</tr>
<tr>
<td><strong>Primary hypotheses (strong):</strong></td>
</tr>
<tr>
<td>P1. Romney, McCain and other Republicans should almost never draw from <strong>FAR LEFT</strong>, and extremely rarely from <strong>PROGRESSIVE</strong>.</td>
</tr>
<tr>
<td>P2. Romney should draw more heavily from the <strong>RIGHT</strong> than Obama in both stages of the 2012 campaign.</td>
</tr>
<tr>
<td><strong>Primary hypotheses (moderate):</strong></td>
</tr>
<tr>
<td>P3. Romney should draw more heavily on words from the <strong>LIBERTARIAN</strong>, <strong>POPULIST</strong>, <strong>RELIGIOUS RIGHT</strong>, and <strong>FAR RIGHT</strong> in the primary compared to the general election. In the general election, Romney should draw more heavily on <strong>CENTER</strong>, <strong>CENTER-RIGHT</strong> and <strong>LEFT</strong> vocabularies.</td>
</tr>
</tbody>
</table>
Baselines

Compare against “simplified” versions of the model:

- **HMM**: traditional HMM without ideological tree structure
- **NoRes**: weaker assumptions (never restart)
- **Mix**: stronger assumptions (always restart)
### Results

<table>
<thead>
<tr>
<th></th>
<th>CLIP</th>
<th>HMM</th>
<th>Mix</th>
<th>NoRes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sanity checks</td>
<td>20/21</td>
<td>19/22</td>
<td>21/22</td>
<td>17/22</td>
</tr>
<tr>
<td>Strong hypotheses</td>
<td>31/34</td>
<td>23/33</td>
<td>28/34</td>
<td>30/34</td>
</tr>
<tr>
<td>Moderate hypotheses</td>
<td>14/17</td>
<td>14/17</td>
<td>12/17</td>
<td>11/17</td>
</tr>
<tr>
<td>Total</td>
<td>65/72</td>
<td>56/72</td>
<td>61/73</td>
<td>58/73</td>
</tr>
</tbody>
</table>
Future and Ongoing Work

- 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)


