Measuring Ideological Proportions in Political Speeches
Supplementary Material

Yanchuan Sim[*]   **Brice D. L. Acree**[†]

[*]Language Technologies Institute
Carnegie Mellon University
Pittsburgh, PA 15213, USA
{ysim,nasmith}@cs.cmu.edu

Justin H. Gross[†]   **Noah A. Smith**[*]

[†]Department of Political Science
University of North Carolina at Chapel Hill
Chapel Hill, NC 27599, USA
{brice.acree,jhgross}@unc.edu

This document contains material to supplement Sim et al. (2013).

A   Topic Labels

The following 61 topics were used to label chapters of books in our ideological corpus: Abortion, Arts, Capitalism, Civil-Liberties, Class, Competition, Congress, Constitution, Courts, Crime, Debt, Defense, Democrats' Lies, Democrats–Failings, Democrats–General, Economy, Education, Elections, Energy, Environment, Family, Federalism, Financial Institutions, Foreign Affairs, Freedom, Future of Nation, GOP Lies, GOP–Failings, GOP–General, Gender, Government Scope, Guns, Health, History (General), History (GW Bush Era), History (Clinton Era), History (Democrats '08), History (GOP '08), History (Obama Era), History (Reagan Era), History (Reagan/Bush Sr.Era), Immigration, Infrastructure, International, Labor, Lobbying, Media, Miscellaneous, Morality, Parties, Patriotism, Personal (Author), Race, Religion, Same-Sex Marriage & Gay Rights, Science, Spending, Tax, Terrorism, Trade, Welfare.

Each magazine was labeled with a document-specific topic. The magazines are: *American Conservative* (Right), *American Prospect* (Progressive Left, *International Socialist Review* (Far Left), *Monthly Review* (Far Left), *Mother Jones* (Progressive Left), *Reason* (Libertarian Right), *The Freeman* (Libertarian Right), *The New American* (Far Right), *The New Republic* (Center-Left), *Z Magazine* (Far Left).

B   Cue Lexicon

Table 1 shows cue terms associated with the ideologies. Figure 1 visualizes the Jaccard similarity between each pair of the SAGE effect vectors.
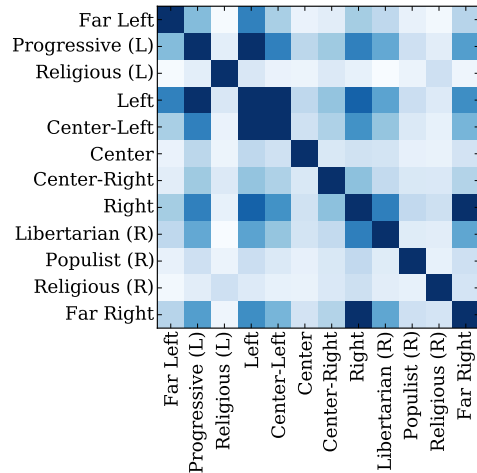

Figure 1: Heatmap showing Jaccard similarity between SAGE ideology vectors. Darker shade implies more overlap between the corresponding SAGE effect vectors. We notice that religious and populist cues are outliers.

Informal Study

As a check of face validity for the vocabularies uncovered by SAGE, we invited several people—four scholars of American politics and three U.S. citizens with a moderate to high interest in contemporary politics—to examine a set of top terms for each fine-grained ideological category. They were given brief descriptions of each class, including prominent prototypical individuals exemplifying each, and asked to match term sets to ideologies. On average, respondents correctly identified about 70% of ideologies using only a handful of terms from each. Experts correctly matched coarse ideologies (LEFT and RIGHT) to appropriate lists of top terms 76 percent of the time (85% LEFT and 70% RIGHT). Several fine (and fairly distinct) ideologies were correctly labeled by all, or nearly all, respon-

dents (LIBERTARIAN, FAR LEFT, RELIGIOUS LEFT and RELIGIOUS RIGHT). Centrist ideologies were sometimes confused but nearly always identified as one of the three centrist categories (CENTER-RIGHT, CENTER-LEFT and CENTER). Of sub-ideologies on the left and right, all but one—POPULIST RIGHT—were given the correct label more than any other label. Relatively few mistakes were made in which LEFT and RIGHT were mixed up.

C   Gibbs Sampler

During the E-step, we perform collapsed Gibbs sampling to approximate the posterior distributions over speaker states, given current settings of the hyperparameters. As the Dirichlet distributions are conjugate priors to the multinomial, we can integrate out the latent variables $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$. For a speech d from epoch e at the ith term, we jointly sample its ideology $s_{d,i}$ and the restart indicator $r_{d,i}$ conditioned on that of all other terms. For simpler notation, we drop the document and epoch level subscripts and denote $\boldsymbol{s}_{-i}$ and $\boldsymbol{r}_{-i}$ as the current state and restart assignments for all other terms except the ith term in document d,

$$
p(s_i = k, r_i = r \mid \boldsymbol{x}_{-i}, \boldsymbol{r}_{-i}, \boldsymbol{w}, \boldsymbol{t}, \alpha, \boldsymbol{\beta}, \rho, \boldsymbol{\zeta})
$$

$$
\propto [(1-\rho)^r (1-(1-\rho))^{(1-r)}]^{\ell_i + 1}
$$

$$
\times \frac{n_{k,w_i}^{-i} + \beta_{k,w_i}}{\sum_{w \in \Sigma} n_{k,w}^{-i} + \beta_{k,w}} \times \zeta_k
$$

$$
\times \left[\prod_{(u,v) \in Path(v_0, k)} \frac{(1-\zeta_u)(f_{u,v}^{-i} + \alpha)}{\sum_{v' \in \mathcal{I}} f_{u,v'}^{-i} + \alpha} \right]^r
$$

$$
\times \left[\prod_{(u,v) \in Path(s_{i-1}, k)} \frac{(1-\zeta_u)(f_{u,v}^{-i} + \alpha)}{\sum_{v' \in \mathcal{I}} f_{u,v'}^{-i} + \alpha} \right]^{1-r}
$$

$$
\times \left[\prod_{(u,v) \in Path(k, s_{i+1})} \frac{(1-\zeta_u)(f_{u,v}^{-i} + \alpha)}{\sum_{v' \in \mathcal{I}} f_{u,v'}^{-i} + \alpha} \right]^{1-r_{i+1}}
$$

where $n_{k,w}^{-i}$ is the number of times word w is generated by ideology k and $f_{u,v}^{-i}$ is the number of times we traversed the edge (u, v) of the ideology tree during epoch e.
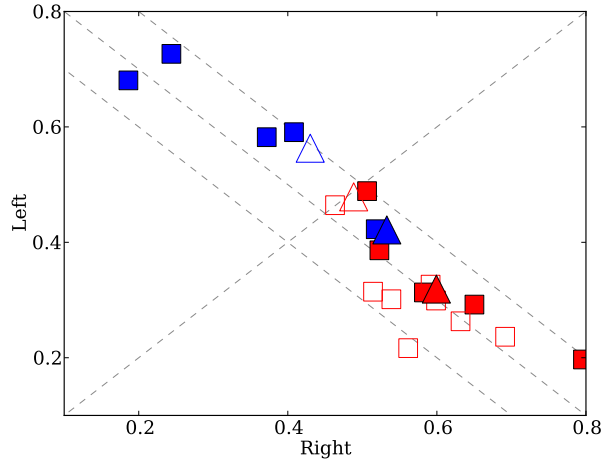

Figure 2: Proportions of time spent in LEFT and RIGHT vocabularies for presidential candidates during the 2008 and 2012 presidential election seasons for CLIP. Republican candidates are denoted in red, while Democratic party candidates are denoted in blue. Solid and hollow points denote the 2008 and 2012 election seasons respectively, while the square and triangle points denote the primary and general elections respectively.

D   Additional Visualization

Figure 2 shows the proportions of time spent in LEFT and RIGHT, aggregating over candidates, by season and party, according to CLIP. Note how general election candidates (circles) tend to be closer to the center. The two blue points leaning RIGHT are Edwards in the 2008 primary election and Obama in the 2012 general election (see the evaluation section for discussion).

References

Yanchuan Sim, Brice Acree, Justin H. Gross, and Noah A. Smith. 2013. Measuring ideological proportions in political speeches. In *Proceedings of EMNLP*.

FAR LEFT (2,802)	monopoli_capit, class_struggl, capitalist_economi, social_movement, occupi_movement, polit_economi, capitalist_system, trade_union, labor_movement, rule_class, develop_countri, world_bank, work_peopl, labor_power, econom_crisi, communist_parti, latin_american, world_economi, latin_america, social_justic fossil_fuel, natur_resourc, nineteenth_centuri, south_africa, work_class, human_develop, multin_corpor, capitalist_societi
PROGRESSIVE LEFT (2,319)	#_peopl, recent_year, abu_ghraib, #_state, execut_director, public_info, state_depart, public_polici, vice_presid, #_centuri, mental_ill, john_kerri, make_sens, polit_parti, presidenti_elect, unit_nation, web_site, interest_rate, york_citi
RELIGIOUS LEFT (941)	biolog_famili, progress_religion, nuclear_famili, mother_teresa, bad_theologi, religi_issu, earli_church, tax_collector, god_love, religi_commun, american_creed, earli_christian, luke_#, church_leader, matthew_#, great_recess, moral_issu, base_initi, evangel_christian, religi_leader, religi_tradit, sundai_school, strict_father, ordinari_radic, god_bless, overcom_poverti, creat_equal, polit_leader, christian_nation, econom_crisi, social_gospel
LEFT (2,580)	north_carolina, econom_polici, execut_director, public_opinion, cell_phone, mental_ill, #_state, abu_ghraib, earli_#, decad_ago, west_bank, presidenti_elect, good_job, air_forc, homeland_secur, polit_system, vast_major, john_kerri, clinton_administr, al_gore
CENTER-LEFT (3,050)	stanlei_kauffmann, modern_art, young_woman, eighteenth_centuri, al_jazeera, nineteenth_centuri, good_deal, young_man, twentieth_centuri, long_ago, mitt_romnei, great_deal, presidenti_campaign, twenti_year, al_qaeda, young_men, john_mccain, hillari_clinton
CENTER (1,230)	long_beach, debt_limit, stock_option, countri_music, averag_american, corpor_america, origin_intent, georg_washington, cousin_john, tax_increas, loan_offic, alexand_hamilton, debt_ceil, park_lot, proof_text, hous_republican, #_visa, hundr_year, jame_madison
CENTER-RIGHT (1,450)	governor_bush, class_voter, health_care, republican_presid, georg_bush, state_polic, move_forward, miss_america, middl_eastern, water_buffalo, fellow_citizen, sam_club, polit_career, american_life, work_class, elect_night, gener_elect, cultur_war, statu_quo, human_digniti, sex_marriag, limit_govern, moder_republican
RIGHT (2,415)	mcmanu_sourc, foreign_aid, north_korea, nation_review, north_american, georg_washington, communist_parti, arm_forc, emphasi_ad, european_union, limit_govern, constitut_convent, presid_georg, presidenti_candid, #_minut, labor_union
LIBERTARIAN (2,268)	medic_marijuana, realiti_taught, intuit_tempt, raw_milk, rand_paul, econom_freedom, health_care, govern_intervent, market_economi, commerc_claus, militari_spend, govern_agenc, due_process, drug_war, govern_polici, minimum_wage, feder_law, #_percent_coalit, percent_coalit, govern_offici, econom_activ, ron_paul, privat_properti
POPULIST RIGHT (1,155)	corpor_america, work_men, border_secur, nation_interest, big_busi, nation_media, birth_rate, hundr_year, special_interest, million_peopl, american_citizen, immigr_law, open_border, mass_immigr, border_patrol, trade_deficit, latin_america, cultur_war, great_nation, elect_offici, forti_year, twentieth_centuri, immigr_reform, white_america, public_servic, work_peopl, american_job, american_compani
RELIGIOUS RIGHT (960)	d&c_#, dai_saint, holi_spirit, matthew_#, john_#, jim_walli, modern_liber, individu_liberti, posit_law, god_word, jesu_christ, elementari_school, natur_law, limit_govern, emerg_church, privat_properti, plan_parenthood, christian_nation, christian_faith, roman_#, foot_soldier, joseph_smith, polit_correct, hundr_year, school_district, psalm_#, cultur_war
FAR RIGHT (2,410)	mcmanu_sourc, ron_paul, north_american, emphasi_ad, american_citizen, foreign_aid, european_union, world_govern, communist_parti, north_korea, constitut_convent, govern_spend, #_centuri, #_amend, polic_offic, central_bank, execut_branch, limit_govern, govern_offici

Table 1: Top cue terms associated with each coarse/fine-grained ideology, with the total number of terms in brackets. The terms are ordered by log-deviation weights in the η^i vectors. # denotes any numeral.