

Research Statement

Brice D. L. Acree¹

Summary My research agenda centers on the development of new tools to analyze large, unstructured data. Much of my work to date, including my dissertation and several ongoing projects, focus on understanding how people—opinion leaders, politicians and average citizens—express their ideological beliefs through the written and spoken word. In two recent papers, and in a chapter of my dissertation, I focus specifically on how candidates for political office use language strategically to send signals about their ideological beliefs to voters. In a set of ongoing projects, I am working with researchers to evaluate how parties and policy ideas are branded by political elites, and how these brands diffuse through news media. In this statement, I will discuss my research agenda, both as a graduate student and as the foundation for my ongoing research as a faculty member.

Ideology & Rhetoric

When candidates or political leaders speak to citizens, they give us a sense of what they believe and how that informs the decisions they would make in office. These beliefs extend far deeper than simple policy preferences. In primary elections, for instance, candidates can largely agree on policy, but differ in how they explain their positions to voters. Do they justify positions mostly on religious grounds? Do they focus most of their speech on certain issues? How harsh or aggressive, congenial or complimentary, is a candidate's rhetoric when speaking about various groups, like minorities, Christians, liberals, or gun owners? Are immigrants 'workers' or 'aliens'? I am interested in studying how politicians construct an ideological brand through their language, and how they tailor these brands toward strategic goals.

Over the past decade, methodologists have introduced new methods for extracting meaningful signals from text. In my first semester of graduate school, I began a project using such computational methods to analyze the debate rhetoric of presidential candidates. This led to a joint project with Justin Gross and his co-PIs, funded by the National Science Foundation, to develop an improved modeling approach for measuring the ideological content of political texts. Current text-analytic methods in political science adopt a bag-of-words representation of political documents, which provides a convenient modeling simplification but also ignores word order. This project, published in the peer-reviewed *Empirical Methods on Natural Language Processing* (EMNLP), combines the computational efficiency of bag-of-words models with the information provided by word ordering in a document.

Our approach, which we call the Cue-Lag model for Ideological Proportions (CLIP), allows researchers to classify the ideology of a speaker based on their language. Inference unfolds in two stages: First, we use a sparse-additive model to identify a vocabulary of multi-word phrases that convey an author's ideological perspective. Then we tailor a hidden Markov model to represent speakers, such as political candidates, as transitioning between latent ideological states manifested through speakers' use of phrases from the vocabulary. In a separate paper presented

¹Departments of Political Science and Statistics, University of North Carolina at Chapel Hill, 300 Hamilton Hall, Chapel Hill, N.C., 27599. Web: <http://www.unc.edu/~bd1acree>, Email: brice.acree@unc.edu

at the Political Methodology, MPSA and APSA annual conferences, we present the CLIP model to a political science audience. We describe how researchers can tailor the model to many substantive understandings of ideology—e.g., unidimensional, hierarchical or categorical—and use model inference to generate ideological scores based on speeches, press releases or other text.

In “Etch-a-Sketching: Testing the Post Primary Moderation Hypothesis,” on which I am the primary author, I use the CLIP model to test for ideological convergence of candidates during presidential general elections. Scholars of American politics expect, nearly as a matter of course, that candidates for office will pursue partisan extremes in order to win primary elections, and then tack toward the political center to win over the general election median voter. Yet such moderation is difficult to observe since candidates also want to avoid ‘flip-flopping’ on issue positions. With coauthors Amber Boydstun, Justin Gross, Yanchuan Sim and Noah Smith, I use the CLIP model to measure the ideological *content* of candidate speeches, rather than scaling a candidate’s issue positions. We find very strong evidence for post-primary moderation in presidential campaigns. The paper serves as my writing sample, has been presented at the MPSA and APSA annual conferences, and is currently under review at the *American Journal of Political Science*.

My dissertation, which I will complete by May 2016, expands on this collaborative research and advances methods for extracting important features from political texts. My first chapter extends the structural modeling approach to supervised topic models. My second chapter explores how to model political text using deep learning computational methods. The third chapter charts the dynamics of political rhetoric, both over the course of presidential terms and over the course of political campaigns. Taken together, my dissertation demonstrates promising new methods for answering substantively-important questions related to political language and ideology. My dissertation includes software written in C++ and callable from R for easy implementation by scholars.

Exploiting Deep Structure in Political Rhetoric Political text presents some unique challenges that make reliable estimation of ideological language difficult. Ideology is a rich theoretical construct, and expressions of ideological beliefs actually comprise many separate, related lexical features. The *topics* that elites speak about carry some information about their ideological beliefs; as do the words they use when discussing those topics, and the general sentiment or tone of their speech about those topics. To properly estimate the ideological content of a text, a researcher must simultaneously account for variation in a document’s topic, language, and tone.

Chapter 1 builds on my past work and contributes to the structural topic modeling approach made popular by Roberts, Stewart, Tingley & Airolidi (2013). With a theoretically-motivated model of ideological rhetoric, I show the advantages of adding additional structure to the traditional logistic-normal topic model, especially for measuring complex language features like ideological perspective. This additional structure comes from three sources: variation in ideological language *over time*; differences in ideological attention to various *topics*, and differences in the *tone* with which ideologues discuss political issues. The first two components are employed widely in political methodology, though almost always separately; and tone—i.e., *sentiment*, or positivity/negativity of language—is not often discussed in political research. This chapter resolves that gap by providing a straightforward model and inference procedure to account for all three sources of information simultaneously.

Chapter 2 explores the use of deep learning techniques for the classification of political language. ANNs are models that attempt to learn patterns in data by passing information through layers of nodes, each of which applies a regression-like weighting function to the input. The structure of multi-layer ANNs allow them to account for high-order interaction and nonlinearities when mapping inputs (e.g., word or phrase use), to outputs (e.g., document labels, speaker ideology, sentiment). This, in turn, can improve classification accuracy in some natural language processing contexts. Deep-learning neural networks have not been discussed broadly in political science methodology, and this article shows how ANNs can serve as an important classification tool in computational text modeling.

The chapter has three methodological sections: 1. constructing multilayer neural networks and estimating model parameters; 2. making inferences on new data; and 3. assessing performance and avoiding common problems like overfitting. I then describe additional possible structures for ANNs, paying particular attention to how word embedding vectors can capture even complex lexical features in text. Word embeddings, made popular by Google's word2vec or Stanford NLP's GloVe, reduce the dimensionality of text data by projecting words into a lower-dimensional space. Distances between words in this space serve as a measure of how similar words or phrases are to each other. In both Monte Carlo and real-data applications, I show that ANNs, achieve competitive classification accuracy compared to other methods. This performance is often improved by including word embedding vectors in the ANN specification.

Dynamics of Political Rhetoric The third chapter of my dissertation explores the dynamics of political rhetoric. Previous work on elite responsiveness to public opinion suggests that leaders should tailor their political messages to prevailing public opinion. Using the methods described in the first two chapters, I measure the ideological content of political rhetoric over time. First, I explore the ideological content of campaign speeches by major candidates in the 2008, 2012 and 2016 elections, and measure the relationship between their language and the dynamics of the campaign. Second, I measure the variation in presidential rhetoric using major speeches and statements over the course of Obama's presidential term, and compare the ideological content of his statements against aggregate public opinion.

I show that over the course of presidential elections, candidates tend to adapt to prevailing opinion of the electorate. During the primaries, candidates tend to maintain fairly partisan language, but they move strategically even within this electoral stage. Barack Obama began his primary campaign using predominantly liberal language. As Obama's popularity and delegate count grew, he transitioned slowly to a more generalist, balanced rhetoric we would expect of a major-party nominee. Meanwhile, Clinton followed an opposite trajectory. At the start of the primary, when Clinton was the prohibitive favorite for the nomination, she employed fairly moderate ideological language. As she lost ground to Obama, however, she began adopting an increasingly liberal voice.

Other Interests

Bayesian Statistics I am interested in Bayesian inference, broadly defined. Bayesian inference can often improve estimation through regularization, smoothing or sampling, and frequently improves the calibration of uncertainty estimation. In a forthcoming paper in the *Journal of Ex-*

perimental Political Science, Timothy Ryan, Matthew Wells and I explore the causal relationship between personality types and how individuals react to disturbing political news coverage. In order to measure the psychological concepts ‘need for affect’ and ‘need for cognition’, we employ a Bayesian latent variable model, which allows for us to incorporate measurement uncertainty in our subsequent analyses. In a working paper, I employ another latent variable approach to estimating political ideology in the public. Following Feldman & Johnston (2009), I cluster respondent preferences into discrete groups. Using Monte Carlo simulation, I consider current approaches to estimating the best number of latent classes to specify *a priori*, and find that most methods underestimate the number of classes. These methods also underestimate the uncertainty inherent in not knowing the proper number of latent classes. I propose a nonparametric approach to mix over the possible number of latent classes in an LCA model, providing better calibrated measures of uncertainty.

Legislative Diffusion An important substantive question in state politics centers on how, when and why states borrow legislative ideas from other states or from interest groups. This poses an interesting methodological challenge: to measure the similarity of highly-technical legislative texts, and to use this measure to trace the diffusion of legislation through the states. Two methods predominate in the state politics literature: cosine distance and longest-substring similarity between documents. Together with three other UNC graduate students, I evaluate these methods with Monte Carlo simulation, and propose a simple distance metric that combines the logic of these two competing measures: a sub-string weighted cosine distance between documents. We use the weighted cosine measure to estimate the influence that interest group model legislation has on policies implemented in the states.

Media As part of a research group I organized in 2015, I have worked with a team of other graduate students to collect and process the transcripts from all available television news programs from 2000 through 2014. The data supports several different research papers, including a project under my direction exploring methods for measuring issue attention and idea diffusion through traditional media. Tracing the evolution in attention over time requires particular care, since scholars need principled ways of measuring when new issues emerge, and when issues disappear from national attention. We are working to measure how political ideas and figures are branded in news media, and how these brands evolve over time.

My research focuses on improving statistical methods for unstructured data. I am a firm believer in collaborative research, and many of my projects include researchers with varied substantive interests and from several different disciplines. I would be happy to discuss these projects with the OSU political science faculty. More information on my research, interest and projects can be found on my website at <http://www.unc.edu/~bd1acree>, and I can be reached by e-mail at brice.acree@unc.edu.