A Discussion on Using Natural Language Processing to Analyze Conservative Think Tank Strategies

---- From a Processual Perspective

Chen Liang

"To continue as things, they (things) must be remade moment to moment." – *Social Process*

Ideologies are often assumed as a fixed set of political views by computational political science, but they are embodied by different actors in different social spacetimes as flexible and, even conflicting, political actions. In this paper, I present a natural language processing method to model the main conservative topic issues in a semantic space, and then discuss its potentials and limitations from a processual perspective.

Introduction

Historically, the conservatism in the U.S. is a mixture of three strands: economic conservatism, traditionalist conservatism, and anti-communism (Busch 2011)[i]. Economic conservatives, or libertarians, promote liberty, small government, and free-market principles. Traditionalist conservatives, or social conservatives, oppose to certain changes in social norms and seek for the preservation of conventional family and religious values. Anti-communist conservatives, with less commitment to domestic issues, have evolved into the so-called foreign policy conservatives within the modern Republican Party, with an emphasis on national defense and national security affairs.

But those who attempt to construct a conservative narrative compatible with all the three strands mentioned above often end up finding themselves either logically inconsistent or overly idealistic. The most distinct and persistent tensions come from libertarians. If individual liberty—freedom of choice and individual judgement—should be maximized, on which ground does the government intervene the sacred private sphere and prohibit something like gay marriage? If capital needs to freely flow across countries to promote comparative advantages, then can the government both promote the economic and technological development of its international competitors and defend its national interest?

Despite the ideological tensions, the legacy of the New Deal—an ever-growing federal government—sparked a conservative coalition between traditionalists and libertarians. As the conservative movement since the late 1950s grew increasingly active, both theoretical and political attempts were made to unite elements of the three strands together. One of the most influential theories was Frank Meyer's philosophy of fusionism, which utilizes "libertarian means in a conservative society for traditionalist ends."[ii] Meyer openly promoted the idea as the senior editor of the conservative journal *National Review,* but his

insistence on limited national power and state autonomy—to the extent that even Abraham Lincoln should be blamed for his "destruction of the autonomy of the states"[iii]—received severe backlashes.

The collective opposition to big government also stimulated the emergence of conservative think tanks. The American Enterprise Institute (AEI) was founded in 1938 "dedicated to defending human dignity, expanding human potential, and building a freer and safer world. The work of our scholars and staff advances ideas rooted in our belief in democracy, free enterprise, American strength and global leadership, solidarity with those at the periphery of our society, and a pluralistic, entrepreneurial culture."[iv] Although widely considered a right-leaning think tank, AEI describes itself as nonpartisan: "AEI takes no institutional positions on policy issues (whether or not they are currently before legislative, executive, or judicial bodies) or on any other issues.[v]" The research Institute, however, has been framing its own agenda to fuse the conservative factions together. The former AEI president, Arthur Brooks, for example, writes that conservatives should be united under a common belief in "conservative social justice." The power of work is "not just a paycheck, but a social network, a source of identity and dignity, an institution that allows people to contribute their talents and derive value from that contribution." Promoting "conservative social justice" demands educational reform and better entrepreneurship in both building business and building one's family and community.[vi]

Nevertheless, this view of conservative social justice is rooted in a classic libertarian belief in promoting entrepreneurship. The theory is ambiguous in terms of how dignity and entrepreneurship should be promoted in private spheres such as family, and therefore still omits most part of religious and moral concerns of the traditionalists. Maybe more profoundly, libertarianism is rooted in a set of science-based (if economics is considered science) studies and therefore fails to provide a satisfiable and *exclusive* framework for the value-based traditionalism. Moreover, most of AEI's research work presents a deeper and more practical concern for conservatives: most of the influential libertarians were academic researchers better at writing lengthy papers than actively advocating with straightforward policy recommendations[vii].

Due to a sense of crises developing from conservative academics "leaving the field (of legislation shaping) to liberal academics and liberal institutions like Brookings,"[viii] the Heritage Foundation was founded in the February of 1973. "Unlike think tanks of the past that published a study and hoped someone would pay attention to it, Heritage strategically regarded its publications as products to be delivered to the right policy makers and news media at the right time." Lee Edwards, a scholar from the Heritage Foundation writes. He also cited Ed Feulner, the president of the Heritage Foundation for about forty years, "Ideas do have consequences, but only if they become part of the public policy dialogue."[ix] Edwin Feulner identified himself as a follower of the Chicago School of Economics and an economic libertarian, but his urge to engage in real-time policy dialogue brought about another way to fuse the

ideological conflicts among conservatives: The more politically feasible policy option is always the better, not matter it is libertarian or traditionalist.[x] With this consequentialist perspective, the Heritage Foundation expanded its influence quickly especially after Ronald Reagan adopted many of his policies based on Heritage's policy study, *Mandate for Leadership*.

But after the relatively closely cooperation between economic and social conservatives during the Reagan Administration, they were forced to compromise as the Bush Administration endorsed the bi-partisan bill to offer states federal K-12 funding[xi] and created the Medicare prescription drug entitlement[xii]. The division among economic conservatives themselves between big-company supporters mostly backed by oil companies and classic free-market, anti-monopoly proponents also deepened the tension. Moreover, the close connection between academic conservatives and political conservatives demands the academic conservatives to not step back but either compromise ideologically or directly give up the persistence in creating a logically consistent and fixed ideological framework. In this process, the traditional conservative coalition dissolves into fragmented factions united for a common interest in electoral success, embracing new members such as blue-collar workers[xiii] and African Americans. The Heritage Foundation, known for its radical traditionalist agenda, named Kay Coles James, a Christian African American female activist famous for her education agenda, as its president in 2017.

Despite this ideological shift, current quantitative measures tend to analyze individuals and organizations on a bi-polar spectrum of left and right. I argue that to defeat political polarization, it is important to capture the nuances in how ideologies are interpreted and framed differently and how these differences themselves lead to political consequences. *I start by discussing how modern Natural Language Processing (NLP) may help us capture the change in conservative think tank strategies. I will then examine the implications and limitations of these approaches from a processual perspective and propose directions for future research.*

Data Collection and Preprocessing

The Heritage Foundation (Heritage) and American Enterprise Institute (AEI) are selected as the research objects for this project under the following considerations. First, both of them are in D.C. and are widely considered influential in federal-level legislation. Second, they tend to focus on a broad range of issues—i.e. not policy-specific think tanks like the Center for Strategic and International Studies (CSIS). Finally, they are independent institutions—i.e. not related to an academic institution (the Hoover Institution, for example, belongs to Stanford University) or a government agency (RAND Corporation, for example, is financed by the United States Armed Force). I plan to expand the research focus to the other conservative think tanks (CSIS, Hoover, RAND, etc.) in the future, but as a primary research, the two think tanks are the one of the best options.

In addition to its own website, Heritage Foundation founded *The Daily Signal*, a conservative political media news website in 2014, with a focus on political commentary. I web-scraped *The Daily Signal* website https://www.dailysignal.com/ as well as the AEI and Heritage Foundation's official site www.aei.org and https://www.heritage.org/. I then conducted text-level and sentence-level data cleaning procedures with the help of Natural Language Toolkit (NLTK) sentence tokenizer.

The cleaned datasets contain 16, 266 articles from Heritage Foundation website (1977-2020), 30, 473 articles from *The Daily Signal* (2008-2020), and 42, 979 articles from AEI (1966-2020). Because the availability of Internet has shifted dramatically in the 21st Century, I focus primarily on the Obama and Trump administration (from 2009 to 2019).

I also obtained a dataset of *Policy Review*, a journal founded by Heritage in 1977 but was acquired by the Hoover Institution in 2001. However, the digital version online contains only 844 articles from 1989 to 2013, and the size of texts before 2001 is further limited. Therefore, I focus my quantitative research on Heritage, the *Daily Signal*, and AEI for this paper. Because Heritage maintains the photocopies of the older publications, I will keep them for future research.

Language Shift

To start with, the shift of language may affect topic modeling in that people may use different words to describe the same things overtime. More generally, as a publication tries to expand its influence, it tends to use more common words for readers with limited education to better comprehend its content. Therefore, I conduct a brief examination on the change of language patterns.
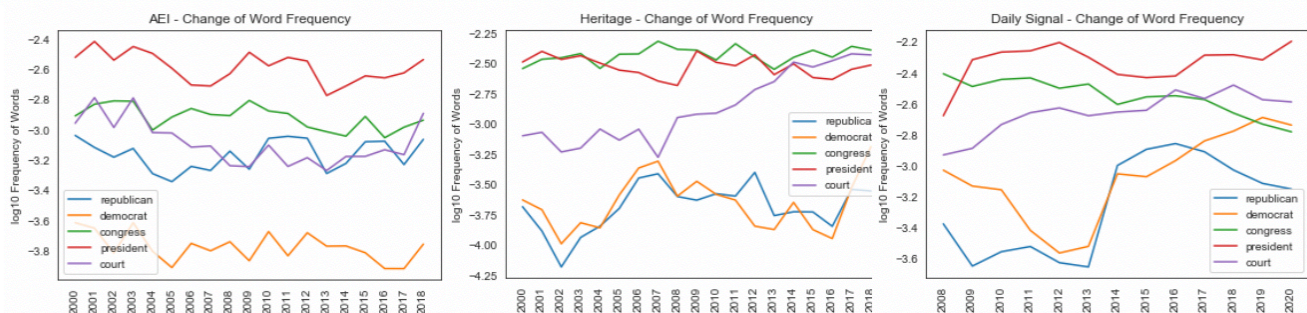
As Michel et al demonstrated based on millions of digitized books in English from 1800 to 2000[xiv], the shift of language can be quantitatively measured from multiple different perspectives, such as the evolution of grammar (irregular words, etc.) and word frequency. Because our observation window is only about one decade, the change of grammar occurs rarely. Thus, I only present the change of word frequency for this project. With a special concern regarding the length of words as they represent a level of democratization in literature, I also examine the change of word length.

I start by calculating the log10 frequency for words *republican, democrat, congress, president,* and *court*, the words which symbolize most of the party-based political dynamics. Most of the frequency trends are fairly stable, but there are also quite interesting patterns. For example, Heritage website and *The Daily Signal* use the words *republican* and *democrat* almost at the same rates, but AEI mentions *republican* far more than *democrat*. But why does the self-perceived relatively neutral think tank, AEI, mention the two parties less balanced than the widely considered radical and partisan think tank, Heritage Foundation? One of the most intuitive explanations is probably that, Heritage cares more about attacking their opponents, i.e. Democrats, than AEI, which focuses more on interpreting the Republican agenda. This intuition is further confirmed, as *The Daily Signal* mentions significantly more *democrat* than *republican* during the Obama's Administration and after 2017, when the campaign to impeach President Trump began.

I then focus on word length by calculating $Index_{year_i} = \sum_{j=7}^{15} j * Prob(len(word) = j \mid year_i)$
I limit the words of interest to those with more than 7 and less than 15 letters to focus on the change of relatively lengthy words. The inclusion of shorter words will dilute the result as they disproportionally dominate the data. In the corpus of interest, words more than 15 letters are rare and mostly typos. The resulting length index is essentially a weighted average word length for each year. Based on this index, we can conclude the following findings.

First, the stereotype is that Heritage is more populist and AEI more elite. This is indeed true for *The Daily Signal*, Heritage Foundation-funded media targeting the general public, which uses significantly shorter words. But Heritage Foundation itself is not quite different from AEI in terms of word length. In fact, AEI has seen a gentle decrease in word length until 2012, probably because of an expansion of its policy interest from economics—which uses many lengthy technical words—to broader conservative dialogues. Moreover, the word length of Heritage surged from 2006 to 2009, probably triggered by anger

and anxiety from Democrats winning control of both houses of Congress in the 2006 U.S. elections and then again from the presidential election in 2008. But it is probably also because of the realization that they were becoming more elite (shown here as using more lengthy words), Heritage started *The Daily Signal,* which target the general public rather than political elites.

As a conclusion, the brief examination of language shift above shows that 1) the frequency and length of words imply differences in think tank writing styles, and 2) the differences are not too significant for us to conduct topic modeling techniques.

Topic Modeling

In this section, I introduce a topic modeling approach to analyze how think tanks place the same policy topics in different semantic spatial positions. I argue that this computational approach is effective in helping us break the shell of ideology and investigate the internal heterogeneity of conservatism interpreted by think tanks.

1. Possible Clustering Methods & Cross-Tag Comparisons

Natural Language Processing (NLP) literature has used both supervised and unsupervised ways to cluster documents, mostly based on either semantic similarities or density distributions. Because my primary interest for this project is to see how documents are distributed over *known* clusters—topics such as economics, religion, national defense, I did not use an unsupervised algorithm. Topic modeling is nevertheless different from traditional supervised classification questions, in that we are interested in not only *whether* a document belongs to a cluster, but *how much* a document belongs to all available clusters. In another word, this is more like a Fuzzy c-means clustering problem.

Topics	Target Words
Economy	Business, Finance, Enterprise, Tax, Market, Trade, Monetary, Tariff, Bank
Religion	Christian, Church, Religion, God, Baptist, Bible, Protestant, Catholic, Religious
Defense	Communist, Defense, Army, Terrorism, Nuclear, Islam, Military, Regime, Troop, Submarine, Iran
Congress	Senator, Congress, Vote, Election
Health	Health, Medicine, Medication, Hospital, Medicare, Medicaid
Woman	Baby, Abortion, Pregnant, Parenthood, Childbearing
Gun	Gun, Firearm, Shootings, Shooting
Tech	Nasa, Cyber, Cybersecurity, Scientist, Hacker
Environment	Environmental, Conservation, Renewable, Energy, Resort, Pollution
LGBT	Homosexual, Gay, LGBT, Sexual, Transgender, LGBTQ
Immigration	Immigrants, Immigration, Border, Mexican, Crossing, Deportation

School	Tenure, School, Tuition, Student, Teacher

I started with a dictionary of words for each topic. These keywords are used to tag documents and calculate the semantic dimensions for topics. I selected the keywords based on a criterion of sufficiency: If any of the selected word appears in a document, this document has to be related to the corresponding topic, and the word should be frequently used in this context. Therefore, ambiguous words such as "security" are not included because it may refer to homeland security or social security. I made this dictionary arbitrary but with some *recursive* adjustments based on Gensim (explained later) word similarity: i.e. if a word is measured as semantically similar to the words in this cluster and it also satisfies the criterion of sufficiency, then I will add the word to its corresponding category. Admittedly, this keyword dictionary is not objective, and it is hard to prove whether the results concluded from such a model is robust. Tuning this dictionary to improve robustness or finding better ways to cluster the document will be a next step of this research.

I then used Word2Vec to train a neural-network model which positions words that share common contexts in the training corpus closer in the space. Word2Vec assumes that words can be presented as vectors such that vector addition and subtraction have semantic meanings: $\overrightarrow{China} - \overrightarrow{Beijing} + \overrightarrow{Tokyo} = \overrightarrow{Japan}$, for example.[xv] For document classification, I calculate the normalized summation of all the keyword vectors in a topic as its topic vector and the normalized summation of all the tokenized word vectors in a document as its document vector.

Admittedly, a more intuitive way to cluster documents is probably to use Doc2Vec, an NLP tool for representing *documents* as vectors. It trains a model called Distributed Memory version of Paragraph Vector (PV-DM) based on the texts and their key-word tags, treating each tag as the topic of the corresponding texts. As a result, at the end of training, each tag holds a numeric representation of all its corresponding texts. But as shown in the graph on the top left, the cosine similarity between each pair of tags is not quite informative. We anticipate that LGBT should be highly similar to religion, and health should be highly similar to women. None of these are observed in this heatmap. A possible explanation is that we tag the documents overly arbitrarily: As long as a document mentions a keyword, we tag the document with the corresponding topic, but the author might just briefly mention the word for a comparison or analogy and then continue on another topic. Our sufficiency criterion, at the very best, can only guarantee that a particular paragraph is about certain topics.

2. Cosine Similarity Heatmaps

The heatmap graphs show the cosine similarity calculated with Word2Vec. The resulting cross-tag similarities largely match with our political intuition. *LGBT* is similar to *religion*; *health* is similar to *woman* (abortion) and *school* (childcare); *environment* is similar to *tech*; etc. The three conservative publications and institutions have similar narratives in most topics: I expect such distinct patterns in liberal think tanks like Brookings.

But there are also some subtle differences and surprises: Heritage ties LGBT much more with religion than the others do—even more than its far-right outreaching branch, *The Daily Signal*. Also, against my
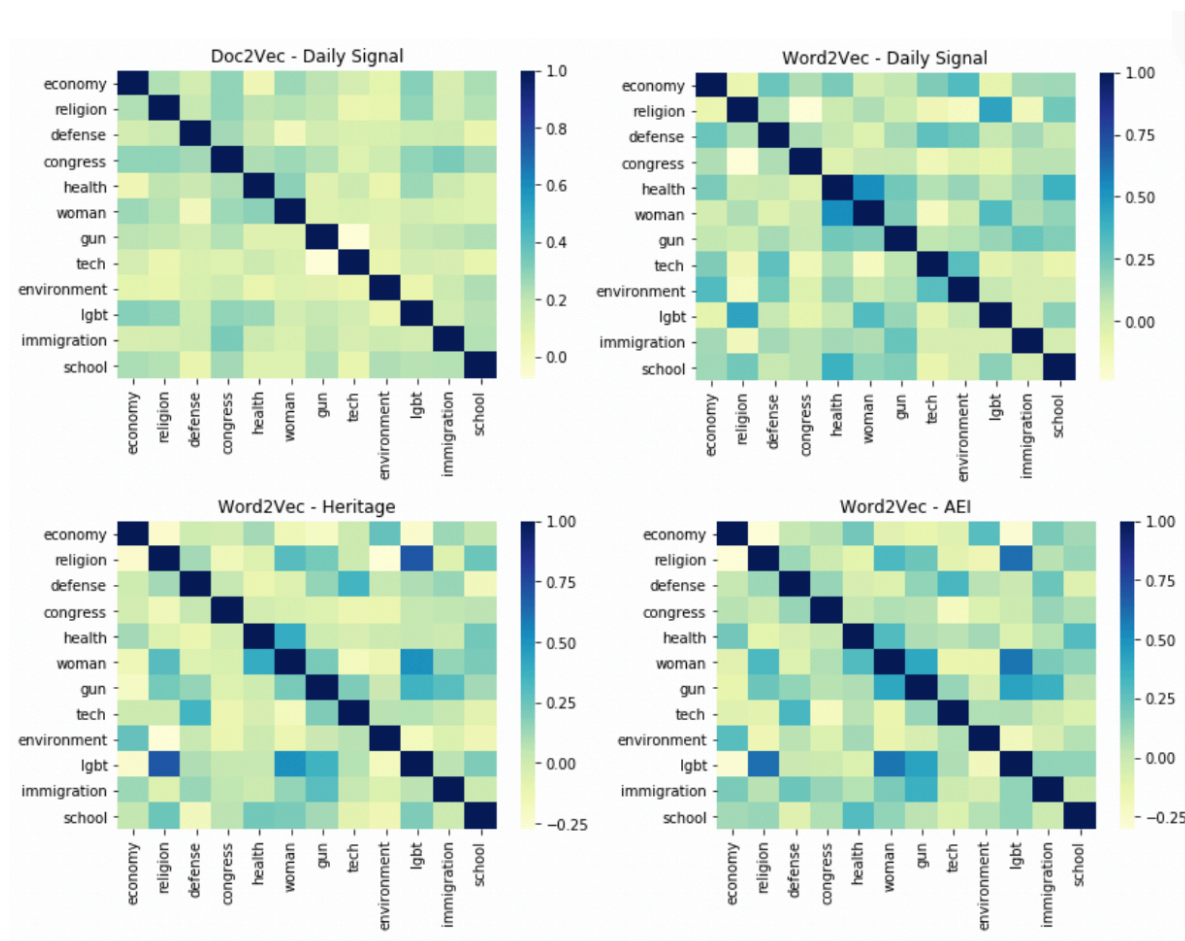
expectation, AEI does not actually tie *defense* that much with *economy*. AEI is known for analyzing the U.S. trade with China from a national defense perspective and advocating for the trade war with China. This is probably because AEI has such a huge national defense research team which focuses also on the Middle East and Russia that, international trade is just one part of its defense research.

3. Topic Visualization with Term Frequency-Inverse Document Frequency (TF-IDF) Vectorizer

Nevertheless, it is still tricky to read and compare among heatmaps. To better demonstrate how topics are positioned in a semantic space, I vectorize the documents using a TF-IDF Vectorizer and then PCA-
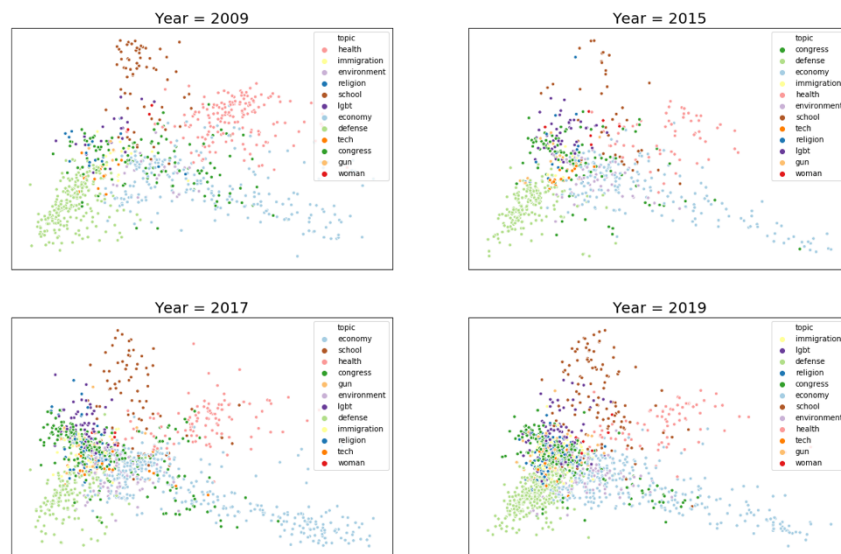
transform the high-dimensional vectors into two-dimensional vectors for visualization. For each document, I calculate its document vector (the summation of all its word vectors) and classify it with the topic which has the highest cosine similarity with this document vector.

I applied the TF-IDF method to vectorize documents without re-using the Word2Vec method, as a way to cross validate the interpretability and robustness of both methods. TF-IDF transforms a document into a N-dimensional vector which represent how important a set of N feature words are to this document in a corpus. Two documents have similar vectors if they use these N feature words with similar frequency distribution. I then use principle component analysis (PCA) to transform the vectors into 2D vectors. As a linear transformation, PCA will not alter the spatial relationship between vectors. In short, I can use the two methods to scatter documents as points on a 2D graph based on their relevancy to each other.

Daily Signal Topics

Year = 2009

Year = 2015

Year = 2017

Year = 2019

Heritage Topics

Year = 2009

Year = 2015

Year = 2017

Year = 2019
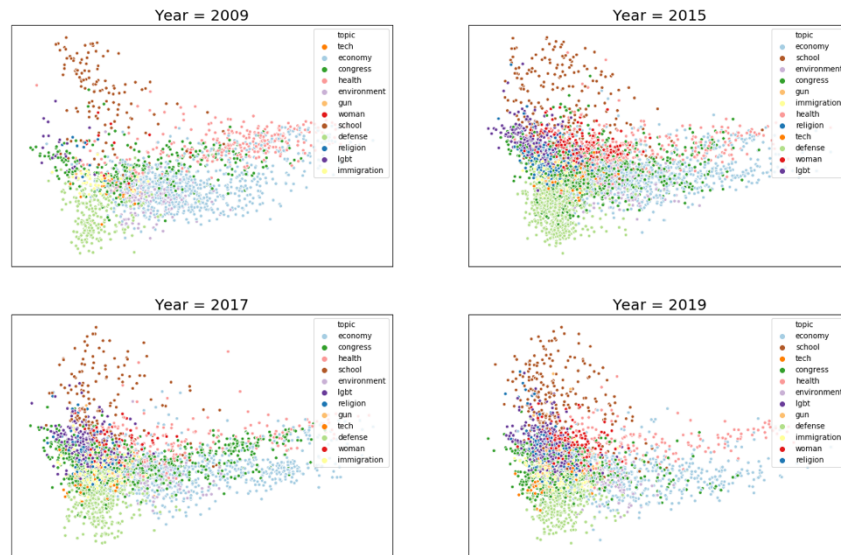
The resulting scatterplots show how policy topics are framed relative to each other. Take *the Daily Signal* as an example. The boomerang-shape distribution indicates three major distinct topics, *defense*, *economy*, and *school*, connected mostly by a number of topics in the middle, namely *woman, congress, religion, lgbt,* and *immigration*. Some of the documents on *economy, health* and *congress* are distributed far away from other topics but closely related to each other. These topics probably involve more technical and political terms not shared by other topics. It also suggests that *the Daily Signal* tends to ground their public health agenda more on economic considerations and legislative negotiations.

The topic distributions for the *Daily Signal* and Heritage are observably different. Unlike *the Daily Signal*, topics like *health* and *congress* are separated from *economy* for Heritage. The word frequency distributions for *health* and *economy* are still distinct from other topics and their separation suggests a subtle division in framing strategy. The research branch—Heritage as a think tank, does not analyze and advocate public-health policies to policymakers much in terms of economic development and welfare. But Heritage broadcasts their public-health agenda to the general public through its outreaching branch—the *Daily Signal*—by connecting the agenda more with financial considerations.
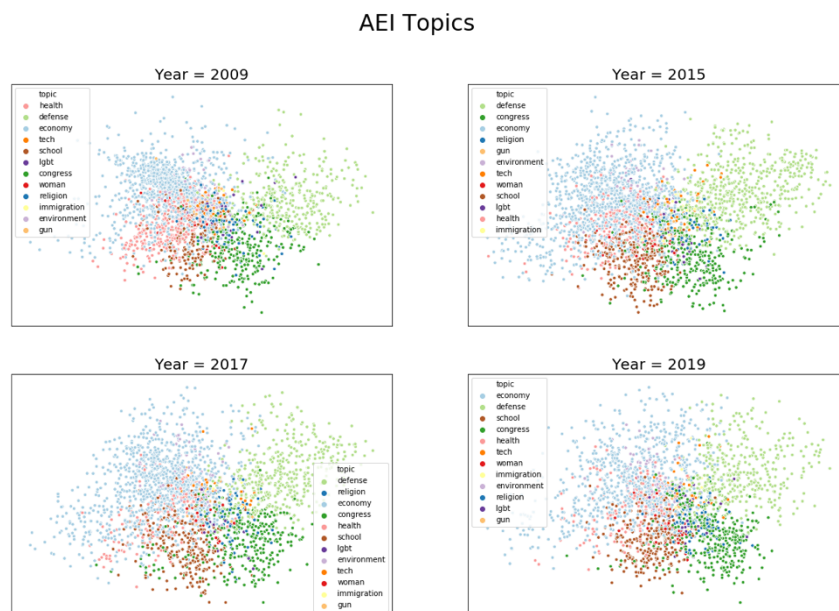
As a comparison, AEI has quite different strategies in framing the same set of topics. It has four distinct major clusters, *economy*, *defense*, *congress*, and *school*, stretching over the four corners of the space with the other topics mixing at the origin. This indicates that the AEI research teams on these four topics are more separated with distinct writing styles and framing strategies.

In 2009, *health* is relatively distinctive as a separated topic of its own, positioned between *economy* and *school*. *Health* is close to *school*, probably because public health problems were analyzed in terms of childcare and student insurance. But it is also possible that AEI, the old-fashioned research institute, cited many articles in their writings and thus mentioned "School of Health Science" and "School of Public Health" more in citations, skewing the semantic meaning of *school* to a different direction from what is expected. However, in 2019, *health* has already blended into *economy*. If we assume that the writing and citation styles are relatively fixed, then AEI at least partially shifts from viewing public health problems as a mixture of social and financial considerations to treating it more financially.

This topic modeling approach also confirms a known distinction between research-based and action-based think tanks. Think tanks are widely considered research institutes which perform research and advocacy on policy topics[xvi], but there has been a debate on whether think tanks should simply conduct unbiased policy research or actively engage in advocacy. AEI claims itself "non-partisan" research institute despite its right-leaning agenda, but Heritage publicly announces that it needs to actively promote conservative policies. This difference is displayed through our topic modeling approach: While Heritage and its publication, the *Daily Signal*, tend to mention congress-related topics when analyzing other issues, AEI organizes its narratives on the congress-related topics more separately. As shown in the scatterplots above, for Heritage and the *Daily Signal*, the congress-related documents center at the middle of the semantic space, connecting all the other topics. But for AEI, the congress-related documents center at the bottom right, presenting only about one fourth of the arguments. The overlaps between *congress* and other topics are also limited. Therefore, it is indeed observable that Heritage tends to analyze other policy topics with a mind to connect these topics with elections and congress votes, while AEI tends to analyze issues case by case without necessarily bringing about domestic politics.

Part-of-Speech (POS) Analysis

Despite the impressive interpretability of topic modeling, another essential part of the story is still missing: What are the perceived actors for these conservative think tanks? Who are the enemies or components, who are the friends, who are the audiences, and who are the stakeholders, and how do these roles change overtime? A simple transition of ideological framework is meaningless in policy research if we fail to understand the transition of political actors who embody the ideologies into political actions. Therefore, the third section of my quantitative research is a tentative and still ongoing project on extracting information from texts to construct connections among actors, actions, and subjects. For this report, I simply introduce the first step of this section: That is, identifying different types of actors.

MIT's SpaCy provides a powerful tool for Named Entity Recognition (NER). The algorithm classifies named entities into pre-defined categories such as organizations, locations, monetary values, etc. Its model has been trained on the OntoNotes 5 corpus and therefore is fairly accurate in capturing at least common entities. NLTK provides a similar model but it needs to be trained to improve accuracy. For simplicity and accuracy, I chose to use SpaCy NER for the following research.

To start with, I use SpaCy to identify political groups, organizations, and laws. SpaCy supports the corresponding entity types: Nationalities or Religious or Political Groups (NORP), Buildings, Airports, Highways, Bridges (FAC), and Companies, Agencies, Institutions (ORG). In practice, SpaCy often misclassifies agencies into location names, which is understandable: The White House is both an agency and a location. Therefore, I combine the classified ORGs and FACs together. Because think tanks rarely

talk about specific locations, when we focus on only high-frequent words, the ORG-FAC approach exhibits satisfiable accuracy.

The harder part is to identify nouns that describe humans. For each noun or pronoun or proper nouns, identified by SpaCy's Part-of-Speech (POS) model, I loop through the first four definitions provided by WordNet, a database of semantic relations and meanings of words. I identify a word as a human-related noun is any of the four definitions contains "someone who," "people who," "person who," "man," and "woman." This method has multiple drawbacks. First, a noun may refer to a human or not based on its context. For example, the second definition of *dog* from WordNet is "a dull unattractive unpleasant girl or woman." Therefore, *dog* is classified as a person-related noun. Second, the method still misses some important words. For example, the only noun definition of *activist* is "a militant reformer." As a result, the word *activist* will not be recognized as a person-related noun. I am working on building a better algorithm to detect person-related nouns, but here I will just use the current method, which still offers some meaningful insights as long as we are concerned of its limitations and biases.

Nationality or Communities

Nationality or Communities, identified as Nationalities or Religious or Political Groups (NORP) by SpaCy, refer to the groups of people mentioned by the think tanks. These groups may represent themselves or the enemies, or simply research subjects. The change of their frequencies tells us how the think tanks prioritize these political groups given their perceived political environment.
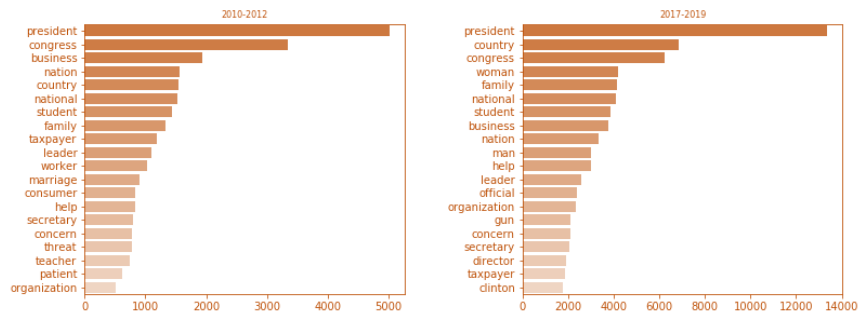
First, the *Daily Signal* and Heritage use proportionately far more *American* and *Americans* than AEI does, echoing the point we made above that Heritage has stronger priority on domestic affairs. In fact, with a little risk of overinterpretation, we may also infer that Heritage takes "American First" more seriously by keeping reinforcing the sense of patriotism.

Second, *Christian* and *conservatives* are frequently used by the *Daily Signal* and Heritage, but not AEI. This further demonstrates that what Heritage has been emphasizing is not only a sense of nationality, but also a Christian-based conservative identity of being an American. Moreover, while AEI covers a wide range of foreign countries with both democratic (Japan, Italy, France, etc.) and authoritarian governments, almost all of the foreign countries Heritage and its publication mention are authoritarian regimes (Russia, China, Iran, North Korea, the Soviet Union, etc.) The difference, again, tells that AEI conducts more comprehensive research on foreign affairs, while Heritage mostly mentions foreign countries as subjects of criticism to reinforce national identity.
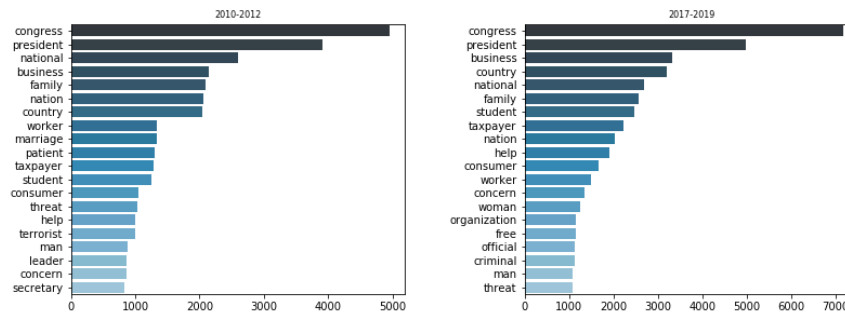
Word Frequency for AEI - Person-related Nouns

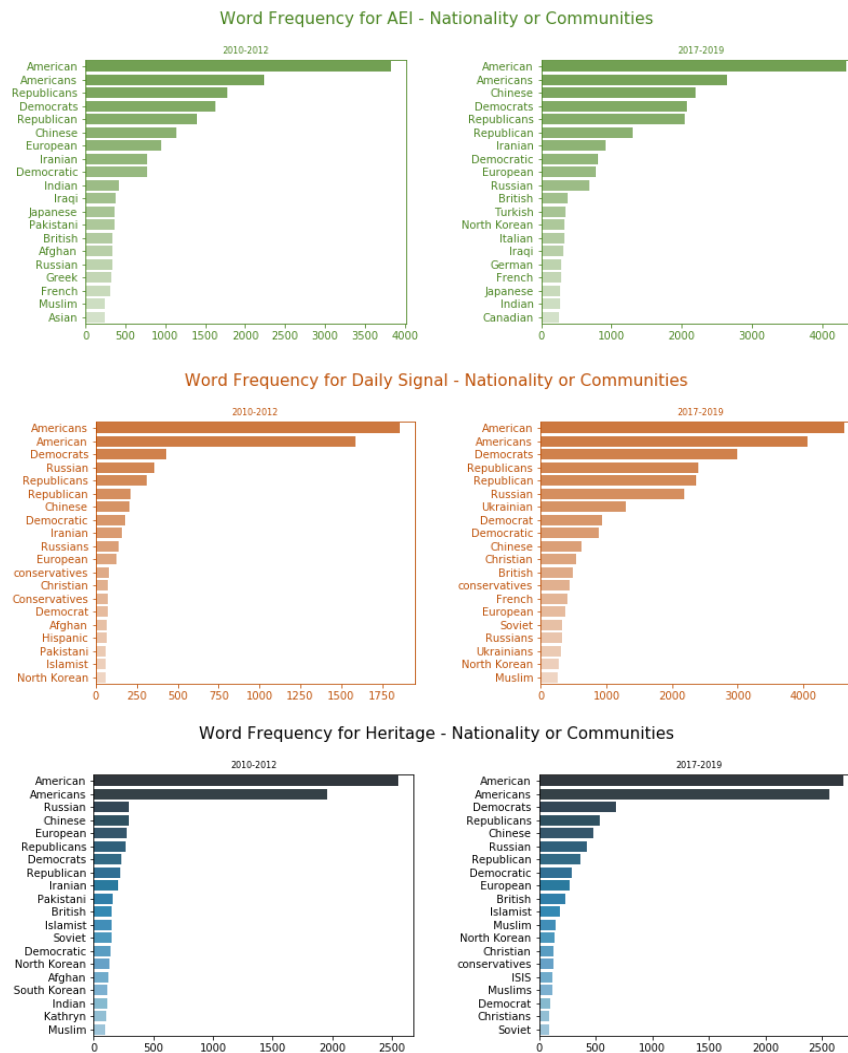2010-2012

president
country
business
congress
national
leader
student
worker
nation
help
man
woman
teacher
consumer
scholar
threat
great
candidate
fellow
resident

2017-2019

country
president
student
worker
business
consumer
help
family
congress
man
national
free
woman
nation
leader
great
economist
model
concern
teacher

Word Frequency for Daily Signal - Person-related Nouns

2010-2012

president
congress
business
nation
country
national
student
family
taxpayer
leader
worker
marriage
consumer
help
secretary
concern
threat
teacher
patient
organization

2017-2019

president
country
congress
woman
family
national
student
business
nation
man
help
leader
official
organization
gun
concern
secretary
director
taxpayer
clinton

Word Frequency for Heritage - Person-related Nouns

2010-2012

congress
president
national
business
family
nation
country
worker
marriage
patient
taxpayer
student
consumer
threat
help
terrorist
man
leader
concern
secretary

2017-2019

congress
president
business
country
national
family
student
taxpayer
nation
help
consumer
worker
concern
woman
organization
free
official
criminal
man
threat

Person-Related Nouns

Word Frequency for AEI - Nationality or Communities

2010-2012

American
Americans
Republicans
Democrats
Republican
Chinese
European
Iranian
Democratic
Indian
Iraqi
Japanese
Pakistani
British
Afghan
Russian
Greek
French
Muslim
Asian

0   500 1000 1500 2000 2500 3000 3500 4000

2017-2019

American
Americans
Chinese
Democrats
Republicans
Republican
Iranian
Democratic
European
Russian
British
Turkish
North Korean
Italian
Iraqi
German
French
Japanese
Indian
Canadian

0   1000   2000   3000   4000

Word Frequency for Daily Signal - Nationality or Communities

2010-2012

Americans
American
Democrats
Russian
Republicans
Republican
Chinese
Democratic
Iranian
Russians
European
conservatives
Christian
Conservatives
Democrat
Afghan
Hispanic
Pakistani
Islamist
North Korean

0   250   500   750 1000 1250 1500 1750

2017-2019

American
Americans
Democrats
Republicans
Republican
Russian
Ukrainian
Democrat
Democratic
Chinese
Christian
British
conservatives
French
European
Soviet
Russians
Ukrainians
North Korean
Muslim

0   1000   2000   3000   4000

Word Frequency for Heritage - Nationality or Communities

2010-2012

American
Americans
Russian
Chinese
European
Republicans
Democrats
Republican
Iranian
Pakistani
British
Islamist
Soviet
Democratic
North Korean
Afghan
South Korean
Indian
Kathryn
Muslim

0   500   1000 1500 2000 2500

2017-2019

American
Americans
Democrats
Republicans
Chinese
Russian
Republican
Democratic
European
British
Islamist
Muslim
North Korean
Christian
conservatives
ISIS
Muslims
Democrat
Christians
Soviet

0   500   1000 1500 2000 2500

As mentioned above, person-related nouns are the nouns with semantic definitions that involve human subjects. Because the semantic meaning of words depends on their context, this classification is not quite accurate. For example, words like *business* (a commercial or industrial enterprise and the people who constitute it), *free* (people who are free), *model* (a person who poses for a photographer or painter or sculptor), and *concern* (a commercial or industrial enterprise and the people who constitute it) are classified as person-related nouns, although they are not commonly used as such. Nevertheless, since we have been using the same method, cross-corpus comparison can still yield meaningful interpretations.

Here we can confirm many of previous conclusions again. First, Heritage mentions proportionately much more *congress* than AEI, indicating that the former may have a clearer mission of reaching out to policymakers and advocating for urgent changes. Second, the topics AEI are mostly concerned about are student-related and business-related policies, which are both cores for libertarianism: Private schools and enterprises should be deregulated. As a comparison, Heritage and its publication, though still caring about deregulation, emphasize more on family and marriage: That is, it emphasizes more on family values and tradition.

Word Frequency for AEI - Organizations and Locations

Word Frequency for Daily Signal - Organizations and Locations

Word Frequency for Heritage - Organizations and Locations
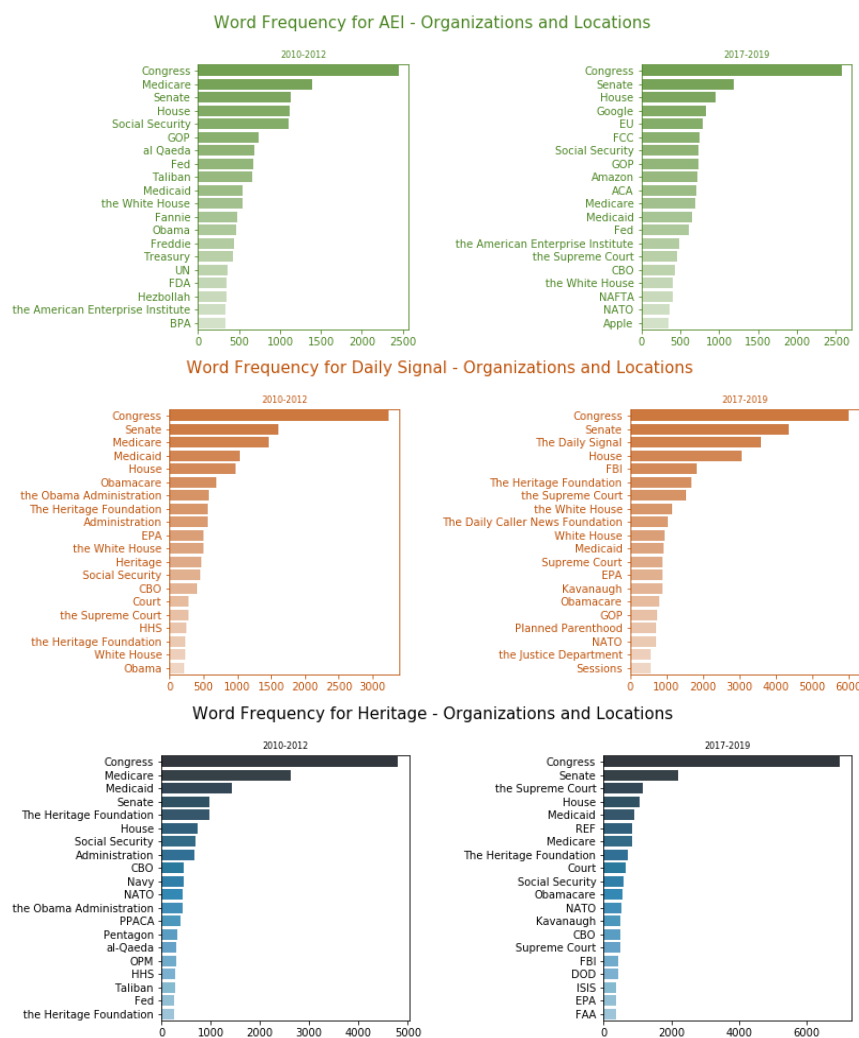
Organizations and Locations

Here we focus primarily on agencies, institutions, and other organizations. First, both think tanks and Heritage's publication maintain a consistent interest in social security and public health programs. However, instead of simply analyzing the policies themselves—Medicare , Medicaid, and the Affordable

Care Act (ACA)—as AEI does, Heritage and its publication pay more attention to the government agencies, particularly United States Health and Human Services (HHS), who are the main actors in shaping public health policies. The action-driven think tank is indeed fulfilling its mission.

Moreover, from 2010-2012 to 2017-2019, AEI has shifted its interest on financial policy (Fed, Fannie, Freddie, Treasury) and terrorism (al Qaeda, Taliban, Hezbollah) to a concentration on technology development, especially net neutrality (Google, FCC, Amazon, Apple). It is harder to conclude what *the Daily Signal* and Heritage Foundation have been focusing on. But compared to AEI, both emphasize more on environment (EPA) and, during the Trump Administration, impeachment (FBI) and the Supreme Court (Kavanaugh), as President Trump nominated Brett Kavanaugh as a Justice of the Supreme Court.

As a conclusion, the quantitative methods introduced above effectively and interpretably show patterns of how the think tanks change their strategies overtime and how they are different from each other. It proves the idea that conservatism is not a well-defined, homogeneous, and fixed concept. Its meaning is flexible and relies on its substance, such as conservative think tanks and their analysts.

Discussion from a Processualist Perspective

The quantitative methods above present a feasible way to measure how conservative think tanks prioritize and change their strategies and agenda overtime. But NLP analysis, especially topic modeling, makes strong underlying assumptions on the durability of the research subjects: the think tanks, language patterns, the semantic meanings of topics, etc. From a sociopolitical perspective, these assumptions are intrinsically questionable, in that definitions and ideas themselves are fluid social constructions. In this section, I examine both the potentials and the limitations of studying the evolution of conservatism based on think tanks, particularly from a processualist perspective.

Academic, non-partisan studies on how think tanks interact with ideology frameworks are rare and often treated as a part of social movement research. A typical challenge of this type of research is that the membership and structures of think tanks, as well as the definition of think tanks itself, are too ambiguous for cross-time and cross-space comparisons to define its research subjects. The tale circling around DuPont Circle in Washington, D.C., where influential think tanks cluster around, says that, despite being conservative, some analysts in AEI deny that Heritage should be considered a think tank at all for being so openly partisan. Moreover, it seems that a think tank may not be independent, e.g. the Stanford-based Hoover Institute and the government-funded RAND corporation. But if so, then the political science and public policy departments in most top universities may also be considered as think tanks with an extreme but logically consistent generalization. The Kennedy School of Public Policy at Harvard and Johns Hopkins University School of Advanced International Studies, for example, have already been playing

important roles in political research and advocacy, with their professors giving testimonies in Congress and serving (or having served) roles in public service.

This ambiguity in definition can take a more complicated form, in that the traditional social thought approaches have to prioritize certain important characteristics of think tanks and omit the others based on their prior assumptions on how think tanks come into being. The tradition of _methodological individualism_ assumes defined and fixed sets of decision rules for individuals, such that collective actions (of a think tank) can evolve through individual choices. But it is hard to examine the roles and preferences of individuals in think tanks. Policy analysts writes on behalf of their supposedly fact-based research, their ideological beliefs, or their previous experience in Capitol Hill, the White House, and the Pentagon, but seldom for their own rank of preferences or for personal decisions as assumed by the individualistic methodology. On the other hand, _conflict theory_ is an overkill. Rarely do think tanks want to overturn an established system as a whole; instead, they play its rules for their own sake.

Methodological sociality instead assumes that think tanks, especially the partisan ones, advocate on behalf of certain classes or ethnic identities, i.e. certain group interests rather than an abstract ideology. This makes more sense: the common donors for conservative think tanks are rich individuals or families, business giants, and religious communities. The stereotype of conservative policy analysts also tends to assume most of them as white, Christian males--which is perhaps not completely wrong.

However, to what extent can we measure think tanks' ideological priorities based on the characteristics of their donors and analysts? The appendix displays the change of major donor foundations (counted as having contributed over 15% of all Heritage funding received in a year) and the Board of Trustee, as well as their brief descriptions. I got the donor foundation information from the Conservative Transparency Project, but the dataset is incomplete: Some of the major donations reported in news are missing from this dataset. The table is thus better read with the Board of Trustee information, because most of the trustees represent their corresponding charity foundations. A stricter due diligence report is definitely needed for more rigorous conclusions, but the general intuition is clear. First, Heritage is mainly funded by finance and financial technology companies, the entertainment industry, and independent conservative writers and thinkers. Second, the influence of the petroleum industry has been fading away and taken place by agriculture-based firms. This might partially explain Heritage's shift from economic policies to traditional values which the rural, agricultural sectors care more about.

There are, however, two layers of challenges for further research. First, it is hard to compile a complete dataset of donation information because private funds often travel through multiple platforms before arriving at the think tanks. For example, the Schwab Charitable Fund is a major donor of Heritage, but its nature as a donor-advised fund indicates that there are more hidden donors behind the scene. In that sense, the publicly available donation information might be heavily skewed. Second, even if we

obtain the full data of donor contributions, causality is ambiguous. It might be true that Heritage is receiving more money from the agriculture industry than from the petroleum industry, but is the radicalization of Heritage a result of this change in donors? Maybe it was Heritage itself which viewed radicalization a worthy political investment or it just happened to recruit a group of far-right political analysts who gear the research focus towards radicalization. Also, the increasingly polarized political environment overall might cause both the change in donation contributions and the change in think tank strategies.

Computational social science approach, on the other hand, is essentially a *symbolistic approach*. As stated in *Social Process*, "the unit of symbolic analysis is ultimately the sign—something that stands for something else… (The signs can be) desired states of affairs." In this sense, the research on how conservative think tanks shape the conservatism agenda can be interpreted as how the symbolic meanings of the word *conservatism* as well as its associated words have been changed, corresponding to the change of writing styles for think tanks. The trace of symbolism can be found in the topic modeling methods I presented above, as I assume that the words have fixed and distinct semantic meanings.

In quantitative text analysis, the assumption that the semantic meaning for a particular word is fixed and unitary, at least for a defined spacetime, is prevailing. This is perhaps an inevitable trend because for any machine learning model to "learn," we need a sample of enough size; that is, we need to aggregate the contexts of word to derive meaningful results but such aggregation in cultural research could be risky. In addition, I have to arbitrarily select a fixed set of keywords so that the result of text classification can be interpretable. A more endogenous method would be using unsupervised classifiers, but current unsupervised topic modeling methods, mostly considered Latent Dirichlet Allocation (LDA), normally fail to yield interpretable clusters. Such unsupervised methods also require a prior knowledge of the number of clusters to be classified. The rationale of either choosing or tuning the number of clusters makes limited sense at least in a political science research setting.

Given the difficulties stated above, processualism could be important in helping both better interpret the NLP results and propose more insightful research plans. From a processualist perspective, modern conservatism in the U.S. can be seen as a momentary confluence of a name *conservatism* and a loose set of texts, individuals, and institutions. As a part of cultural values, it is a self-sustainable set of events embodied in social entities such as think tanks and personal entities such as policy analysts. Rather than stating that conservatism is a fixed symbol without any meaning of its own, processualism would argue that the shared ideology and interest are more like the momentum which connects Zeno's arrow at each time together. Zeno's arrow, in this context, is the policy position at each social time-space, at which think tanks make decisions based on given constraints and facilities. The constraints and facilities are

defined by their research capacity, financial resources, political networks, as well as the general public opinions, legislative contexts, and policy stakeholders.

In another word, the challenges listed above can be resolved in a processualist perspective, or more specifically, an action-based research. The unit of analysis should neither be individuals in the think tank nor its ideology itself, but the political and social events the entity as a whole is involved: bills they support, testimonies they make, research reports they draft, talks and conferences they hold, etc. At each given thick present, think tanks make decisions based on their knowledge of the past and their perceptions about the future, subject to the constraints and facilities mentioned above. The past consists of their perceived political facts: the threats from the political opponents, the interpretation of religious texts, the Constitution, the other classics they embrace, and the public opinion. The future is shaped by their contemporary definition about what conservatism should be about: That is, the values they hold in the present. Therefore, it is not the individuals, the ideology, or the donors that lead to the change in think tank strategies. Instead, the conservative movement is a set of successive decisions made at each present, observed through transition of individuals and personalities, reinterpretation of ideologies, and succession of donation contribution.

Further, this processualist perceptive explains a reason of political polarization (at least in think tanks). Because both liberal and conservative think tanks are making decisions based on their own perceived past and future, they do not share the same thick present. Even though it seems that they are experiencing the same Newtonian time and Euclidean space, they live in different realities, shaped by their distinct experience in the past and values about the future. Even if they start from the same reality, the cumulation of differences in path dependency will ultimately drive the two parties to completely different routes. Since it is almost impossible to see beyond one's local knowledge and to think in terms of the other reality, a growing trend of political polarization is awaiting.

But this action-based theory also shows us an optimistic possibility. The other methodologies—methodological individualism, etc.—attempt to establish causality over a long period of time. The premise, however, is that current political stances are predominantly bounded by prior causes: personal interests, class and identity conflicts, or certain symbolic-abstract idea construction. If "the grand outside event was held to be inevitable," as *Social Process* writes, then "it would determine all the little events within." Think tank decision makers who themselves believe in such a premise thus lose their potential agency embedded in the social process. The past and future are constraints and facilities, but not the decisions themselves. Actors at the present get to decide a better way to go for the sake of their own communities and the general public.

The final question, however, is how to embody this processual perspective into quantitative academic research, if possible. One potential way, as what I mainly try to do in this paper, is to loosen our

prior definition of issue topics as much as possible. Rather than focus too much on the pattern of ideological words such as *conservatism, republican, liberty,* and *religious*, I place these words in a general semantic space. What I mainly investigate instead, is the issue subjects (gun, student, parenthood, nuclear, etc.) and the political actors (Americans, EPA, FBI, etc.) involved in each present, loosely estimated as each year. Essentially, I tried to understand what the stakeholders and issue subjects think tanks care most about and how they prioritize them accordingly, at each approximate present.

There might be better ways to do it as we expand the horizon and obtain more information. Ultimately, we might end up getting to case studies on how think tanks set their agenda in the face of new crises and emergencies. At the very least, I argue in this paper that, computational text analysis can be used for understanding think tank ideologies from a processualist perspective and, the idea of social process also keeps us aware of the limitations in our current quantitative approaches.

Appendix: Heritage Donation Contributions
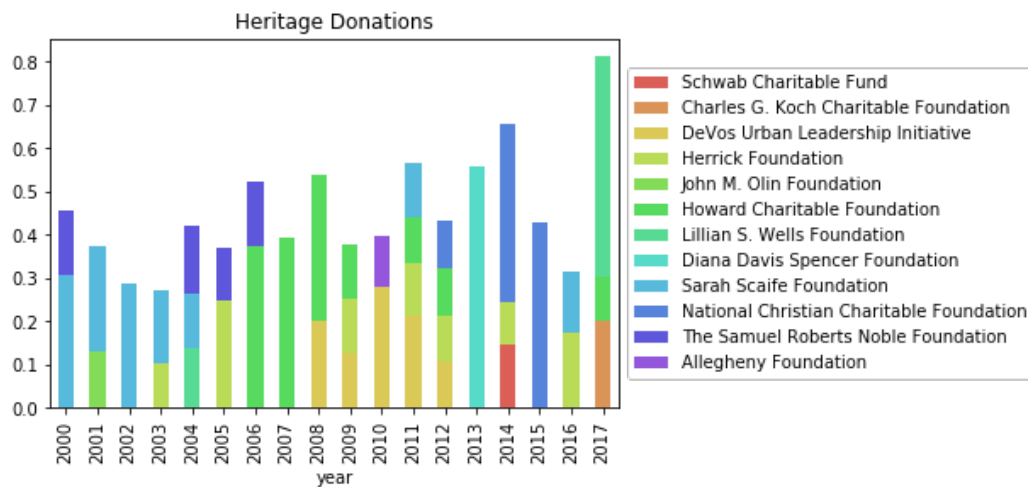
Data Source: "The Heritage Foundation," Conservative Transparency. Data retrieved May 28, 2016. Accessed through https://www.desmogblog.com/heritage-foundation.

Donation Foundations	Description
Schwab Charitable Fund	A Donor-Advised Fund funded by Charles Schwab Corporation, American multinational financial services company
Charles G. Koch Charitable Foundation	The Koch family is a key funder of the right-wing infrastructure, who owns Koch Industries, American multinational oil and refining corporation[xvii]

DeVos Urban Leadership Initiative	A national faith-based leadership development program for urban youth workers in Christian organizations and institutions. The founders own Amway Corporation, American multi-level marketing company that sells health, beauty, and home care products.[xviii]
Herrick Foundation	The Michigan's Herrick family are the former owners of the Tecumseh Products Company. It donates mostly to Michigan-based organizations, focusing primarily on Christian, higher education, and right-of-center advocacy cause.
John M. Olin Foundation	A conservative grant-making foundation disbanded in 2005, initiated to encourage "the thoughtful study of the connections between economic and political freedoms, and the cultural heritage that sustains them." The founder was the president of Olin Industries, chemical and munitions manufacturing company[xix]
Howard Charitable Foundation	A private foundation that primarily gives grants to healthcare and education nonprofits in San Diego, California
Lillian S. Wells Foundation	Failed to find its description, but it seems to prioritize arts and health[xx]
Diana Davis Spencer Foundation	Shelby, an author and investment banker, was the longest serving American ambassador to Switzerland. His wife, Kathryn Wasserman Davis, was an author, community leader, and renowned lecturer in Russian affairs. [xxi]
Sarah Scaife Foundation	late right-wing billionaire Richard Mellon Scaife, whose wealth was inherited from the Mellon industrial, oil, aluminum and banking fortune
National Christian Charitable Foundation	US non-profit organization that assists donors in donating to charitable causes. [xxii]
The Samuel Roberts Noble Foundation	Continue Lloyd Noble's vision to address agricultural challenges and continue this legacy by making charitable grants to nonprofits that cultivate good health, support education and build stronger communities, primarily in Oklahoma[xxiii]
Allegheny Foundation	A right-leaning grantmaking foundation founded by Richard Mellon Scaife. It confines most of its grant awards to programs for historic preservation, civic development and education[xxiv]

Unless further cited, the information above is collected from https://www.influencewatch.org/.

Board of Trustee Information (2012 versus 2016)

Names marked as blue are unique for 2012 Board of Trustee.

Names marked as green are unique for 2016 Board of Trustee.

2012 Board of Trustee

Author collected from the Heritage Foundation website and Exxonsecrets.org.

[i] Busch, A. E. Social Conservatives and Economic Conservatives. *Society*, 49(1), 13–23. 2011, doi: 10.1007/s12115-011-9498-4

[ii] E. J. Dionne, Jr., *Why Americans Hate Politics*, New York: Simon & Schuster, 1991, p. 161.

[iii] Neary, Robert K., and Frank S. Meyer. "In Defense of Freedom: A Conservative Credo." *The Western Political Quarterly 16*, no. 2 (1963): 508. https://doi.org/10.2307/444985

[iv] AEI's Organization and Purposes. *American Enterprise Institute*, www.aei.org/aeis-organization-and-purposes-2/. Date accessed, March 23, 2020.

[v] ibid

[vi] Wassink, Brad. "A Conservative Social Justice Agenda." *American Enterprise Institute*, 11 July 2014, www.aei.org/society-and-culture/a-conservative-social-justice-agenda/. Date accessed, March 23, 2020.

[vii] Edwards, Lee. *Leading the Way: the Story of Ed Feulner and the Heritage Foundation*. Crown Forum, 2013.

[viii] ibid

[ix] ibid

[x] ibid

[xi] No Child Left Behind Act of 2002

[xii] Medicare Prescription Drug, Improvement, and Modernization Act of 2003

[xiii] Lieberman, Ben. "Save the Earth, Sacrifice American Workers?" *The Heritage Foundation*, 9 June 2008, www.heritage.org/environment/commentary/save-the-earth-sacrifice-american-workers. Date accessed, March 23, 2020.

[xiv] Michel, J.-B., et al. "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science*, vol. 331, no. 6014, 2010, pp. 176–182., doi:10.1126/science.1199644.

[xv] " Fiorini, Rodolfo A. "From Computing with Numbers to Computing with Numeric Words." *2017 IEEE 16th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC),* 2017, doi:10.1109/icci-cc.2017.8109734.

[xvi] Diane Stone 'Think Tanks and Policy Analysis', in Frank Fischer, Gerald J. Miller. & Mara S. Sidney (eds.) Handbook of Public Policy Analysis: Theory, Methods, and Politics, New York, Marcel Dekker Inc. 2006: 149–157

[xvii] "Koch Family Foundations." SourceWatch, www.sourcewatch.org/index.php/Koch_Family_Foundations. Date accessed, March 23, 2020

[xviii] "Home: DeVos Urban Leadership Initiative." Home | DeVos Urban Leadership Initiative, www.dvuli.org/. Date accessed, March 23, 2020

[xix] Jmof.org Archived March 5, 2005, at the Wayback Machine

[xx] Barnett, Kari. "Lillian S. Wells Foundation Gives $5M to Top off Parker Playhouse Campaign." *Sun*, South Florida Sun-Sentinel, 15 Oct. 2019, www.sun-sentinel.com/features/society/fl-fea-broward-donation-tops-off-parker-playhouse-20191015-vg3z5kxg2jg7ngdfbulqgiezda-story.html. Date accessed, March 23, 2020

[xxi] "About Us." *The Diana Davis Spencer Foundation,* ddsfoundation.org/about-us/. Date accessed, March 23, 2020

[xxii] "About Us." *National Christian Foundation*, www.ncfgiving.com/about/. Date accessed, March 23, 2020

[xxiii] "The Samuel Roberts Noble Foundation." *Noble Foundation*, noblefoundation.org/. Date accessed, March 23, 2020

[xxiv] "Scaife Foundations.*" Allegheny Foundation*, www.scaife.com/alleghen.html. Date accessed, March 23, 2020