Tracking Theoretical Shifts over Time: A NLP Analysis of Cass R. Sunstein

Omar Cherif[1] Dalila Ladli[1] Maxime Méloux[1] Supervisor: Prof. Samuel Ferey[2]

[1]IDMC, Université de Lorraine [2]BETA, Université de Lorraine

Context

Behavioral economics

In the 1960s, the psychologists Amos Tversky and Daniel Kahneman developed the field of *behavioral economics*. Clashing with the previous assumption from classical economics that humans always act perfectly rationally (*homo economicus*), behavioral economics introduced the concepts of biases, heuristics and bounded rationality[2]. One notable example is *prospect theory*, which relies on the idea that humans assign non-linear values to losses and gains[4].

Cass R. Sunstein

Cass R. Sunstein (born 1954) is an American professor of law, and one of the most cited scholars of the 20th century. He adheres to the Republican thought, a school of political theory revolving around the concept of citizenship in a state organized as a republic. At the beginning of the 2000s, he developed the concept of *libertarian paternalism*, a conception of politics that relies on behavioral economics[5].

Hypotheses

We wish to use Natural Language Processing techniques on Sunstein's articles to investigate the three following questions:

- Did Sunstein's thought transition from republicanism to behavioral economics? When?
- How are subtopics of republicanism (citizenship, civic virtue...) treated over time?
- How are global topics (poverty, autonomy...) treated over time?

Corpus acquisition

We retrieved articles publishing by Sunstein between 1987 and 2005 from two websites: JSTOR (`http://www.jstor.org/`) and Chicago Unbound (`http://chicagounbound.uchicago.edu/`). After filtering duplicates and selecting the highest-quality version of articles, we obtained 231 PDF files. We then used several optical character recognition (OCR) tools to extract the text from those articles, which were often scans.

Corpus cleaning

Manual cleaning

This first stage consisted in removing page breaks, headers, footers, tables and figures from articles, as well as adding missing pages.

In this respect the distribution of powers was consonant with Madison's own hostility to rapid change in government, captured in his antipathy to "turbulence,'" but disso- 930 BRIGHAM YOUNG UNIVERSITY LAW REVIEW [1987 nant with Jefferson's belief that turbulence is healthy for a republic.	In this respect the distribution of powers was consonant with Madison's own hostility to rapid change in government, captured in his antipathy to "turbulence,'" but dissonant with Jefferson's belief that turbulence is healthy for a republic.

Figure 1. An example of manual cleaning.

Automatic cleaning

In a second stage, we wrote a custom program to fix hyphenation, typographical and OCR errors. After this stage, our corpus contains 2.8 million words and 17.5 million characters.

Judicial control of regulatory behavior has come in the form of the "hardlook" doctrine, initi- ally developed by the United States Court of Appeals for the District of Columbia CIrCUIr, 2 and subsequently endosred by the Supreme Court.3 3 The hard-look doctrine has both procedural and substantive elements. Procedurally, it requires regulatory agencies to generate detailed explanations for their decisions-to consider reasonable alternatives.'"	Judicial control of regulatory behavior has come in the form of the "hard-look" doctrine, initially developed by the United States Court of Appeals for the District of Columbia Circuit, and subsequently endorsed by the Supreme Court. The hard-look doctrine has both procedural and substantive elements. Procedurally, it requires regulatory agencies to generate detailed explanations for their decisions — to consider reasonable alternatives.

Figure 2. An example of automatic cleaning.

Statistical analysis: TF-IDF

The first supervised approach consisted in measuring the term frequency - inverse document frequency (TF-IDF) of selected terms over time. The TF-IDF measures the frequency of a word in a document (here, all articles of a single word) relatively to the whole corpus.

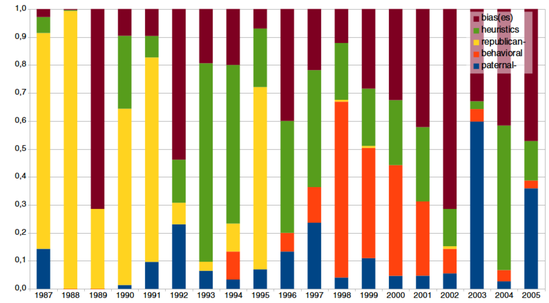

Figure 3. The normalized TF-IDF of selected words over time. We notice a clear shift from *republican-* to *behavioral* and *paternal-*, seemingly occurring in 1996.

Transition modeling: The sigmoid

To model transitions in Sunstein's vocabulary, we fit relative frequencies of words to a sigmoid function with four learnable parameters: $f(L, k, x, x_0, b) = \frac{L}{1+e^{-k(x-x_0)}} + b$. We select words fitting the curve with $r^2 > 0.4$ and categorize them.

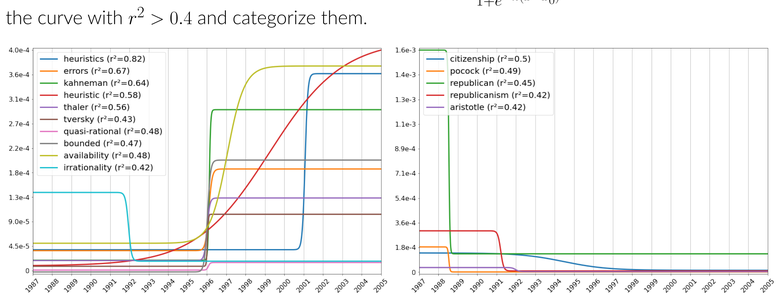

Figure 4. Modeling the transition for two categories of words: behavioral economics (left) and republicanism (right). There seems to be a clear transition around 1996 for the former, while the latter almost disappears.

Semantic analysis: Tropes

Using Tropes (`http://www.tropes.fr/`), a semantic analysis software, we extracted the co-occurrence graph of the Sunstein corpus before and after 1994 and visualized it with Gephi (`http://gephi.org/`).

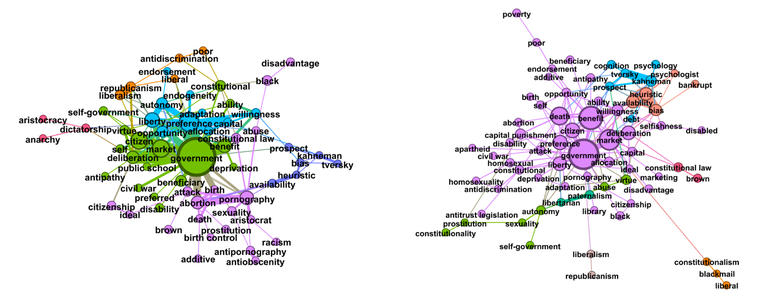

Figure 5. The full word graph before and after 1994. The placement and connections of specific words such as *poor*, *deliberation*, *preferences* and *heuristics* tend to validate our hypotheses. The graph layout was generated using the Force Atlas 2 algorithm and color-coded by modularity clustering.

Topic Modeling

Topic modeling is a field of NLP focusing on extracting main ideas and themes from documents. Topics are lists of words with associated weights: the higher the weight, the more important a word is in a topic. We began by using Latent Dirichlet Allocation [1], a method relying on contextual co-occurrences to create topics.

1987 (topics 1 and 2)	
.011*power	.017*state
.008*administrative	.014*court
.008*judicial	.012*statute
.008*understanding	.010*regulatory
.008*contract	.010*private
.006*state	.009*democracy
.006*agency	.009*protection
.006*justice	.009*provide
.006*lochner	.009*discrimination
.006*system	.009*agency

1995 (topics 1 and 2)	
.020*right	.009*lawyer
.011*people	.007*theory
.009*rule	.007*social
.008*category	.006*people
.008*constitutional	.006*constitutional
.007*political	.006*political
.007*particular	.006*general
.006*might	.005*right
.006*think	.005*rule
.006*change	.005*agreement

2005 (topics 1 and 2)	
.009*would	.014*precautionary
.008*people	.013*principle
.008*paper	.010*discount
.007*activity	.008*risk
.006*agency	.007*regulation
.006*perhaps	.007*supra
.006*engage	.006*benefit
.006*relatively	.020*sunstein
.006*procedure	.011*market
.006*likely	.011*posner

Table 1. The first two topics extracted with LDA in different years. The results are hard to interpret due to a lack of coherence and consistency (topics are re-generated every year).

Bertopic

BERTopic[3] improves on LDA by being able to analyze topic evolution (the same topics are kept through time) and by internally using word embeddings. BERTopic found 160 topics on the corpus, which we grouped into categories. We then analyze their evolution over time.

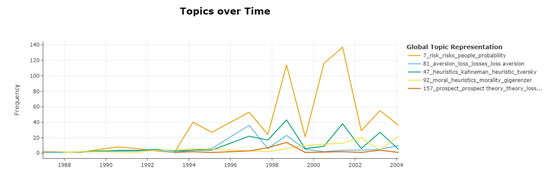

Figure 6. Topics from the "behavioral economics" category first appear in 1994, suggesting a shift around that time.

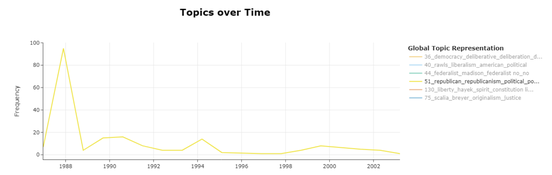

Figure 7. The "republicanism" topic peaks in 1988 but survives until 1994, after which it almost fully disappears.

Conclusion

We have applied four different NLP methods, both unsupervised and supervised, to the Sunstein corpus. All of them seem to indicate that a shift occurred in Sunstein's thought between 1994 and 1996, representing a changeover from republicanism to behavioral economics. This transition can also be detected more subtly in the subtopics of republicanism, such as citizenship and civic virtue, as well as in other global questions, such as that of poverty and deliberation.

References

[1] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3(null):993–1022, March 2003.

[2] Samuel Ferey. *Une histoire de l'analyse économique du droit: calcul rationnel et interprétation du droit*. Number 1 in Collection Droit et économie. Bruylant, Bruxelles, 2009.

[3] Maarten Grootendorst. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*, 2022.

[4] Daniel Kahneman and Amos Tversky. Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–291, 1979.

[5] Richard H. Thaler and Cass R. Sunstein. Libertarian paternalism. *American Economic Review*, 93(2):175–179, May 2003.

{cherif25u, ladli2u, meloux1u}@etu.univ-lorraine.fr