

Building Semantic Networks of Moral Content in News

There is mounting evidence that humans possess innate moral sensibilities, which enable the understanding and enforcement of norms regarding socially-acceptable behavior. A well-known conceptual framework supporting this view is Moral Foundations Theory (MFT). Haidt and colleagues have developed a program of research that separates moral concepts into distinct domains associated with evolved adaptations for living in social groups (e.g., Haidt & Graham, 2007; Graham et al., 2011). They argue that there are at least five underlying domains which serve as the foundation for specific moral rules (Haidt & Joseph, 2004; Haidt & Graham 2007): care, fairness, loyalty, authority, and purity. More recently, the same research group has added liberty as a potential sixth domain (Iyer, Koleva, Graham, Ditto, & Haidt, 2012). Individuals vary in the relative salience of these domains both across and within cultures, and that variation has been used to explain political differences (Haidt & Graham, 2007).

To better understand how morality is communicated through mass media, our research group has developed the News Narrative Analyzer (NNA; [URL redacted for submission]), an open-access resource which integrates automated data collection and analysis alongside input from human coders in order to reliably identify moral content in the news. For both practical and principled reasons, the NNA interfaces with existing free and open-source software to implement features whenever possible, and we have built our own equally-free tools as necessary. The system aims to identify the morally-relevant entities in a given article, map their allegiance or opposition to other entities, and classify those entities - as well as the content of the article as a whole - into the fundamental moral domains theorized by MFT. We hope to use this data to plot how moral content in the news fluctuates in response to major events, explore how political figures are portrayed in moralistic terms, examine how different news sources display proclivities toward certain moral domains, and understand the extent to which the reciprocal relationship between audience selective exposure to and media emphasis on morally-relevant

content impacts attitudes and beliefs.

The NNA uses automated text analysis to provide a preliminary classification of the moral domains referenced in an article. One simple approach to this problem is word-frequency analysis: using a dictionary that maps certain words to specific categories, find what proportion of the words in a document belong to each category. This is the basic logic employed by LIWC (<http://liwc.net/howliwcworks.php>), a nonfree text analysis program widely used in the social sciences. While this approach has inherent limitations, and we are actively pursuing more advanced techniques as we iteratively develop the system, we nonetheless believe that interesting insights can emerge from word-frequency analysis in conjunction with the other data we have collected.

In order to integrate word-frequency analysis into our data processing pipeline, the Python Natural Language Toolkit (NLTK; nltk.org) is used in conjunction with the morality dictionary provided by the MFT research group (<http://moralfoundations.org/othermaterials>). A purpose-built Python script cleans the data (e.g., excluding articles with fewer than 20 words), tokenizes the text, applies the Porter stemming algorithm (Porter, 1980), and then counts the number of word stems associated with each domain-valence category as defined by the MFT dictionary. The valence and primary moral domain referenced by the article is thus classified using word count data.

However, initial results led us to question the validity of the existing dictionary. Articles classified as positively-valenced references to fairness vastly exceeded all other valence-domain combinations in our data set, accounting for nearly twice as many articles as the second-highest category (positively-valenced authority). Of course, it is possible that this reflects the reality of our data. On the other hand, it may reflect a systematic problem. Most articles in our database had very low word-counts in every morality category, with the top category often receiving only 1 or 2 hits, which suggests that the MFT dictionary may have been too limited to accurately capture the variation in our data.

Therefore, we have expanded the MFT dictionary by taking each existing word as a seed to find other semantically-similar terms through a vector-space approach. The free, open-source utility word2vec (<https://code.google.com/p/word2vec/>) is packaged with vector representations of words from publicly-available datasets, such as the Google News corpus. Using word2vec, we found word-vectors that are oriented similarly to each seed word from the MFT dictionary (see [authors redacted for submission], under review). By using this technique and varying the cosine-similarity threshold, we have developed multiple candidates to be used as an expanded morality dictionary. We are actively working to find the optimal similarity threshold to maximize variability in our data without introducing excess noise, validate our expanded dictionary against expert content-analysis and previously-validated experimental stimuli, and release an improved morality dictionary to other researchers. If accepted to the preconference, we will present the results of our efforts to explore this semantic network of moral concepts.

Additionally, the NNA interfaces with the Stanford Named Entity Recognizer (NER; <http://nlp.stanford.edu/software/CRF-NER.shtml>) to find the key entities referenced in each article. We then present each entity as a potential coding unit to trained coders, who select the most morally-relevant entities, classify how the morality of those entities is portrayed in the article, and identify factions of entities who are allied with or opposed to each other. The result is a network of political actors which can be graphed based on their co-occurrence across articles, their allegiance or opposition to each other, or the moral domains they are associated with in the news. We are eager to present initial results regarding the structure of these networks and how they differ across sources at the preconference. We have not yet publicly demonstrated the NNA, and would welcome the opportunity to learn from and share our resources with other researchers in this area.

References

- Graham, J., Nosek, B.A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology, 101*, 366-385.
- Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research, 20*, 98-116.
- Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus, 133*(4), 55–66.
- Iyer, R., Koleva, S., Graham, J., Ditto, P., & Haidt, J. (2012). Understanding libertarian morality: The psychological dispositions of self-identified libertarians. *PLOS One, 7*, e42366.
- Porter, M.F. (1980) An algorithm for suffix stripping. *Program, 14*(3), 130–137.
- [authors redacted for submission]. (under review). Automatic morality annotation using seed words and limited expert labels.