# An Empirical Exploration of Moral Foundations Theory in Partisan News Sources

**Dean Fulgoni**[1], **Jordan Carpenter**[2], **Lyle Ungar**[1,2], **Daniel Preoţiuc-Pietro**[1,2]
[1]Computer and Information Science, 3330 Walnut St.,Philadelphia, PA 19104
[2]Positive Psychology Center, 3701 Market St., Philadelphia, PA 19104
University of Pennsylvania
dfulgoni@seas.upenn.edu, jordac@sas.upenn.edu, ungar@cis.upenn.edu, danielpr@sas.upenn.edu

**Abstract**

News sources frame issues in different ways in order to appeal or control the perception of their readers. We present a large scale study of news articles from partisan sources in the US across a variety of different issues. We first highlight that differences between sides exist by predicting the political leaning of articles of unseen political bias. Framing can be driven by different types of morality that each group values. We emphasize differences in framing of different news building on the moral foundations theory quantified using hand crafted lexicons. Our results show that partisan sources frame political issues differently both in terms of words usage and through the moral foundations they relate to.

**Keywords:** Text Mining, Document Classification, Text Categorisation

## 1. Introduction

Framing is a central concept in political science and journalism (Goffman, 1974). When studying news, each source frames the news within a particular viewpoint, aiming for a change in the perception of the issue among the readers (Scheufele, 1999). Although political and cultural beliefs may be formed and developed through reason and careful thought, it is more common that they stem from automatic, emotional judgments (Haidt, 2001). However, these immediate responses often differ across people; the same action may inspire feelings of disgust, happiness, and anger in different observers. These contrasting emotional reactions lead to drastically different assessments of the acceptability or appropriateness of the action. For instance, the issue of police violence against minorities can be thought of as a conflict of values: liberals abhor the unfairness of violence instigated against a marginalized social group, while conservatives hate the anti-authoritarian, public disrespect of legitimate authority figures.

We perform a large scale analysis of news from partisan news sources across a wide variety of issues. To date, political orientation prediction has been focused mostly on predicting the leaning of individual users (Rao et al., 2010; Pennacchiotti and Popescu, 2011; Al Zamal et al., 2012; Cohen and Ruths, 2013; Volkova et al., 2014; Volkova et al., 2015; Sylwester and Purver, 2015), mostly on social media. While recently, computational approaches have been used to study framing (Tsur et al., 2015; Baumer et al., 2015; Card et al., 2015), little empirical work has been performed to study deeper psychological factors that cause this linguistic divergence (Yano et al., 2010) and analyses have mostly focused on data from political actors (Nguyen et al., 2015).

Social psychology research suggests that differences in the frames used follows the different values that the groups adhere to, and that these values span different issues. Moral Foundations Theory (Haidt and Joseph, 2004; Haidt and Graham, 2007; Graham et al., 2009) was developed to model and explain these differences. Under this theory, there are a small number of basic moral values that people intuitively support, emerging out of both cultural and evolutionary factors, and individuals differ in the relative level to which they endorse these values. The five moral foundations are: care/harm, fairness/cheating, ingroup loyalty/betrayal, authority/subversion and purity/degradation.

It is expected that each partisan source would thus frame an issue to appeal to their reader's moral foundations. Although recent research (Feinberg and Willer, 2015) brings evidence that a better method to convince others of your viewpoint is to appeal to *their* moral values, media usually functions as an echo chamber where partisan sources are made mostly to appeal to people sharing similar values.

In this paper, we aim to shed further light on this phenomenon by performing an analysis of word usage in partisan news sources. We study 17 different issues, ranging in topic from climate change to common core and from abortion to police violence. We show the divergence of themes between the partisan sides, where each side uses different frames to appeal to their readers. We first demonstrate that we can use general word occurrence statistics to identify the partisanship of unseen individual news articles on the same issue. Then, using a manually crafted dictionary, we aim to highlight how often each moral foundation is invoked by each partisan group. The differences are mostly consistent to the moral foundations theory and are consistent to some extent across issues.

## 2. Data

We retrieve data using the Media Cloud API.[1] Media Cloud is an open source, open data platform that collects tens of thousands of online sources, from news to blogs, from across the globe. The news sources are grouped into 'collections' by attributes such as location, type of media, and political partisanship. Further, sets of news stories are grouped into 'controversies' or issues having a common topic e.g. the Isla Vista shooting. Due to legal issues, the end user is only allowed to retrieve word counts for a specific query, not the full text.

[1]http://mediacloud.org/api/

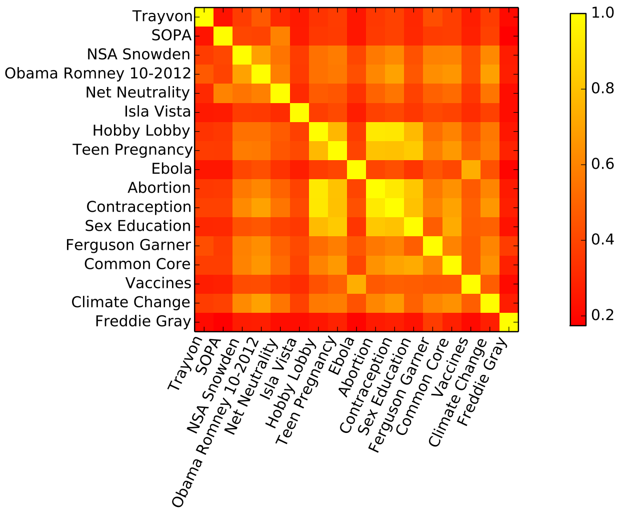



Figure 1: Cosine similarity between issues.

We download word counts from news stories categorised as belonging to a partisan news source to one of three sides in the US: conservative, liberal and libertarian. The news stores are only retrieved if they are part of one of 17 'controversies' or issues, spanning a wide range of topics such as 'Ferguson', 'abortion', 'climate change' or 'net neutrality'. For each issue and partisan side, we identified the top 1,000 unique words, stopwords excluded, and mapped these to their respective counts. These word counts are smoothed for consistency: if any word did not appear, that word was attributed the minimum value of that group.

## 3. Differences in Language Use

As a first exploratory step, we analysed the similarity in word usage across issues, disregarding partisanship. Figure 1 shows cosine similarity between the word frequency distributions for each issue.

Probably the most notable feature of these findings is a cluster of sex-related issues ('teen pregnancy', 'abortion', 'contraception', 'sex education') which seem to all be discussed in similar ways. Surprisingly, three issues that seem related, the violence against Eric Garner, Freddie Grey, and Trayvon Martin, do not contain very similar word distributions.

Focusing on a selected set of issues, we show in Figure 2 the different words each partisan side (limited to liberal vs. conservative for clarity) uses when mentioning a story. These word clouds were generated using the word counts for each partisan bias . For each issue, its respective conservative and liberal word counts were compared using the log odds ratio to determine the relative use of the most polarizing words. Log odds ratio is a widely used technique that performs well in identifying the features most associated with particular labels - in this case, the words dominantly used by conservatives or liberals.

The word clouds consist of 80 words each, the top 40 most conservative and top 40 most liberal words by log odds ratio. Larger text size indicates a higher total frequency for that word. Because some words are used much more than others, the relative frequencies were scaled by taking their logarithm. Red identifies the word as polarized conservative, while blue identifies the word as polarized liberal. Deeper color indicates a stronger correlation with that partisan bias. Thus deep red words are strongly conservative while deep blues ones are strongly liberal.

The patterns of word usage are largely face valid, containing specific aspects of issues of special concern to liberals and conservatives respectively. In general, most partisan sides appear to be focusing on the opposing side and combating their views. Words such as 'conservative' and 'republican' appear highly used in all issues on the liberal side, while 'Obama' and 'democrats' are usually used by conservatives. Also, several word clouds feature each side mentioning contentious and famous opposing voices: conservatives mention 'Gore', 'Pelosi', and 'Sharpton', while liberals mention 'Koch', 'Huckabee', and 'Paul', and even 'Fox' which is the name of the media station known for its conservative stance. For abortion, conservatives use 'pro' and liberals 'anti', which leads to the conclusion that they focus more on combating the opposing sides' positions. Also noteworthy is the use of 'democratic' for liberal sources, which usually appears in the context of 'democratic party', while the conservatives use the word 'democrats' contrasting the difference between using the official name to the word 'democrats' which hints non-affiliation.

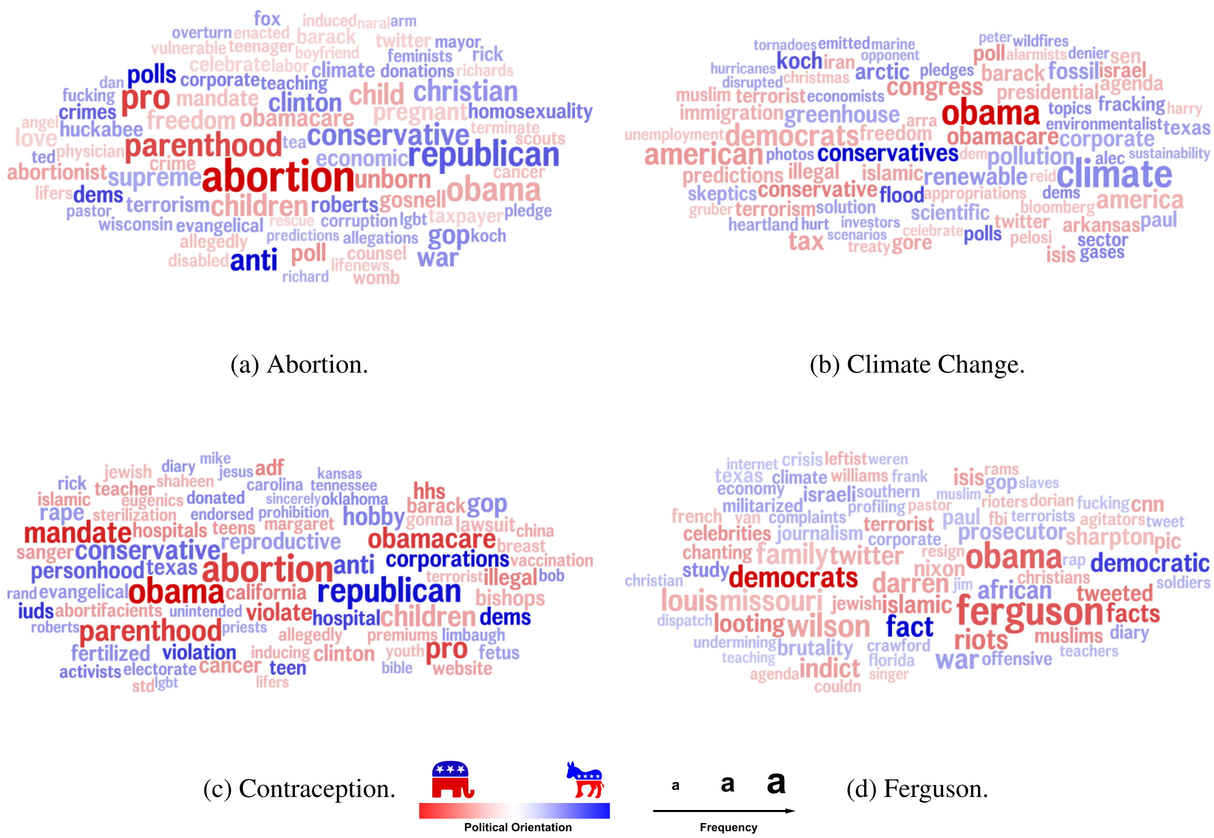
In some topics, there are clearly different frames in use. For abortion, conservatives discuss more about the concrete aspects of the abortion process such as 'children', 'unborn' 'womb', 'pregnant'. Liberals mention concrete issues about climate change (e.g. 'fracking', 'renewable', 'pollution', 'arctic', 'flood', 'fossil'), while conservatives seem to discuss more about their political opposition (e.g. 'congress', 'tax') and apparently other unrelated issues. Conservatives refer to the events involving Ferguson as 'riots', while de liberal side denounce the 'brutality' or 'profiling'.

## 4. Moral Foundations

The Moral Foundations theory is based on the existence of five moral values:

- **Care/Harm**: The valuation of compassion, kindness, and warmth, and the derogation of malice, deliberate injury, and inflicting suffering;
- **Fairness/Cheating**: The endorsement of equality, reciprocity, egalitarianism, and universal rights;
- **Ingroup loyalty/Betrayal**: Valuing patriotism and special treatment for one's own ingroup;
- **Authority/Subversion**: The valuation of extant or traditional hierarchies or social structures and leadership roles;
- **Purity/Degradation**: Disapproval of dirtiness, unholiness, and impurity.

Under this theory, observing an action that undermines a value inspires moral disapproval from feelings of anger, disgust, or contempt (Rozin et al., 1999), while observing an action that supports a value results in happiness, respect, or elevation (Haidt, 2000). For example, a person who strongly endorses the value of *harm* will be appalled at an




(a) Abortion. (b) Climate Change.

(c) Contraception. (d) Ferguson.

Figure 2: The word clouds show the top 40 words most used by each of the two partisan sides (liberal vs. conservative) according to the log odds ratio. Darker red indicates more conservative, darker blue indicates more liberal. Larger size indicates higher word frequency (log-scaled).

action that causes suffering, while someone who endorses *authority* will champion an action that supports the social hierarchy. These responses would be immediate, emotional, and intuitive.

Moral Foundations theory explains some political disagreements between people on opposite sides of the political spectrum. Liberals tend to strongly endorse the values of *care* and *fairness*, but they consider *ingroup loyalty*, *authority*, and *purity* to be irrelevant to morality. On the other hand, conservatives tend to endorse all five moral foundations, though they do not support *care* and *fairness* as strongly as liberals do (Graham et al., 2009). Therefore, some of the most contentious political conflicts center on issues where each side focuses on a value not equally shared by the other side.

In their study of the Moral Foundations Theory, Haidt and Graham (2007) develop a dictionary which is meant to capture people's spontaneous moral reactions. For each of the five moral foundations, this dictionary contains two categories of words: recognition of the value being demonstrated (virtue, denoted with **+**) and recognition of the value being violated (vice, denoted with **-**). For instance, the 'Harm+' sub-dictionary contains words describing compassion and care: 'safety', 'protection', 'shelter', etc. Likewise, the 'Authority–' sub-dictionary contains words that describe rebellion: 'heretic', 'riot', 'nonconformist', etc. The sub-dictionaries contain an average of $\mu = 31.8$ words or stems.

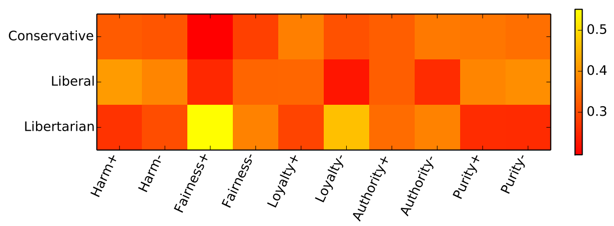
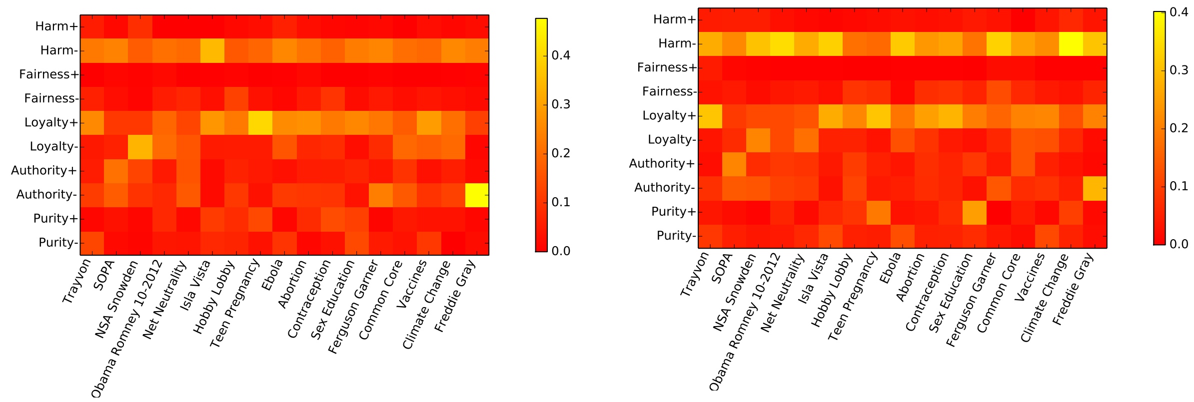



Figure 4: Moral Foundations for Partisan Sources across all Issues. Cells present relative frequencies.

Therefore, in theory, the extent to which a news article uses the words in each sub-dictionary captures its author's concern for each moral foundation. Additionally, we represent each issue and partisan side in terms of the moral foundations they express through text. The most frequent words from each subdictionary in our dataset across all issues are presented in Table 2.

We examine the usage of moral foundations for each of the three side across all issues in Figure 4.

In conformance to theory, liberal sources are far more likely to mention harm and suffering across issues, while conservative sources are more likely to address *loyalty*. The most striking is the use of 'Fairness**+**' and 'Loyalty**-**' for libertarians. Libertarians' patterns of moral values have been found to differ from those of both liberals and conservatives (Iyer




(a) Moral Foundations for Partisan Conservative Sources.

(b) Moral Foundations for Partisan Liberal Sources.

Figure 3: Moral foundation dictionary usage by issue and partisan side.

| Moral Foundation | Virtue (+) | Vice (–) |
|---|---|---|
| Care / Harm | -.194 | -.222 |
| Fairness | -.182 | -.156 |
| Ingroup Loyalty | .098 | .313 |
| Authority | -.002 | .345 |
| Sanctity | -.057 | -.111 |

Table 1: Relative differences in moral foundation word usage across the two partisan sides. Negative values are more used by liberal sources, positive values indicate higher conservative usage.

et al., 2012). From our analysis, libertarians sources put high emphasis on *fairness*, but have the lowest scores on the *harm* dimension.

Figure 4 somewhat lacks interpretability as some moral foundation words are more frequently mentioned overall compared to others. In Table 1 we focus only on the liberal – conservative dimension and show the relative differences between the usages divided by their overall usage. This table highlights that liberals are higher in using words related to the *harm* and *fairness* foundations, just as theory describes. In addition, both the virtue and vice aspects of these foundations are more often invoked. On the other hand, conservatives score higher in the *loyalty* and *authority* moral foundations, similar to theory. However, differently from liberals who focus on both virtue and vice of their foundations, conservatives highly emphasize the vice aspect i.e. betrayal of loyalty and the subversion of authority. The *purity* moral foundation shows the lowest relative differences and, in general, the lowest usage. The direction of usage is inverse to theory, which posits that conservatives value this foundation more. We expect this is either due to this foundation not being explicitly mentioned or the dictionary not being able to adequately capture it. This is hinted by the top words for the 'Purity+' category (e.g. 'church', 'virgin', 'clean') which are not words likely to be used to express appreciation for purity in a political context.

Figure 3 shows in full detail the moral foundations invoked by conservatives and liberals and on each issue. In general, the most used foundations across all issues are 'Harm–' and 'Loyalty+'. However, liberal sources used 'Harm–' significantly more than conservatives across all issues.

Of particular interest is the cell for 'Authority–', which, for the issue of Freddie Grey, is strong for liberals and even stronger for conservatives. This difference probably stems from a particular word within the 'Authority–' subdictionary: 'riot'. Liberal sources were more likely to refrain from using this word, preferring the more sympathetic 'uprising', which is not included in the Moral Foundations dictionary.

Another difference regarding usage of different foundations is the high incidence of 'Loyalty–' for conservatives in the 'NSA Snowden' issue. Indeed, conservatives sources frame Snowden as being a 'traitor' or 'disloyal' to his country, while liberals frame the story in terms of harm caused to the country. On the issue of climate change, 'Harm–' is the most prevalent for liberals across all issues, while conservative sources use 'Harm–' relatively very little and frame the issue more in terms of *loyalty*.

## 5. Predicting Partisanship

To empirically test the differences in framing of different issues, we predict the political partisanship of unseen articles. We choose the Naïve Bayes classifier because we only have access to word counts from the Media Cloud API. The original posts are not available from the Media Cloud platform due to copyright issues. In order to build a testing dataset, we downloaded URLs from the platform for each issue having more than 80 documents, their determined issues and mapping to partisan side. We then crawl the links to obtain the full text of the articles. We experiment with using as training data only the issue-specific statistics as well as overall word statistics across all issues, in order to study if there are any overall terms or frames used by a specific partisan side. Table 3 presents prediction results in terms of accuracy, precision, recall and F1-score for issue-specific training set, as well as baseline accuracy (predicting major-

| Authority Virtue (+) | Fairness Virtue (+) | Harm Virtue (+) | Ingroup Virtue (+) | Purity Virtue (+) |
|---|---|---|---|---|
| law | rights | care | nation* | church* |
| control | justice | secur* | unite* | virgin |
| legal* | fair | protect* | group | clean* |
| order* | equal* | defen* | family | innocent |
| leader* | balance* | safe* | member | pure* |
| **Authority Vice (–)** | **Fairness Vice (–)** | **Harm Vice (–)** | **Ingroup Vice (–)** | **Purity Vice (–)** |
| protest | discriminat* | war | foreign* | sin |
| illegal* | bias* | kill | individual* | disease* |
| refuse | disproportion* | attack* | immigra* | sick* |
| oppose | unfair* | fight* | terroris* | exploit |
| riot* | injust* | killed | enem* | dirt* |

Table 2: Most frequent words from each moral foundation dictionary category from our dataset. '*' denotes a wildcard matching all suffixes.

| Issue | | | | Issue Specific Training | | | | All Issues |
|---|---|---|---|---|---|---|---|---|
| | #Lib | #Con | Baseline | Accuracy | Precision | Recall | F1 | Accuracy |
| Abortion | 328 | 373 | 0.532 | 0.646 | 0.644 | 0.748 | 0.692 | 0.699 |
| Climate Change | 307 | 155 | 0.665 | 0.677 | 0.523 | 0.439 | 0.477 | 0.697 |
| Common Core | 36 | 94 | 0.723 | 0.777 | 0.865 | 0.819 | 0.842 | 0.646 |
| Contraception | 106 | 60 | 0.639 | 0.705 | 0.577 | 0.683 | 0.626 | 0.693 |
| Ebola | 157 | 150 | 0.511 | 0.567 | 0.546 | 0.667 | 0.601 | 0.590 |
| Ferguson Garner | 305 | 411 | 0.574 | 0.644 | 0.657 | 0.793 | 0.719 | 0.619 |
| Freddie Gray | 47 | 48 | 0.505 | 0.684 | 0.641 | 0.854 | 0.732 | 0.632 |
| Hobby Lobby | 123 | 61 | 0.668 | 0.723 | 0.561 | 0.754 | 0.643 | 0.663 |
| Net Neutrality | 55 | 40 | 0.579 | 0.695 | 0.628 | 0.675 | 0.651 | 0.779 |
| NSA Snowden | 190 | 98 | 0.660 | 0.590 | 0.443 | 0.796 | 0.569 | 0.615 |
| Obama Romney 10 2012 | 528 | 407 | 0.565 | 0.585 | 0.518 | 0.668 | 0.584 | 0.591 |
| Trayvon | 45 | 35 | 0.563 | 0.600 | 0.531 | 0.743 | 0.619 | 0.675 |
| Vaccines | 129 | 72 | 0.642 | 0.627 | 0.486 | 0.75 | 0.590 | 0.672 |

Table 3: Prediction accuracy for unseen partisan articles.

ity class) and accuracy when using training data combined across issues.

Our results show that overall, we can predict the partisan side of an unseen article above chance. With a couple of exceptions, perhaps caused by limited data, we are able to improve on the majority baseline up to $\sim 18\%$ accuracy. The best relative improvement is obtained for the 'Freddy Gray' issue, while the highest overall precision is obtained for the 'Common Core' issue. Comparing to using all data in training, there is no consistent pattern: we notice this help in some cases (7 out of 13), while in others it actually hinders performance (4 out of 13) or does not offer significantly different results (2 out of 13).

In order to study in detail this general problem of domain adaptation, we use all train-test dataset combinations to figure out which issues to combine in order to improve predictive performance. Figure 5 displays a heatmap of the relative improvements added by using a different issue in training a classifier for a given issue.

We notice that overall, adding data from other issues hinders the general performance of prediction. There are some exceptions to this, as issues which are related help improve in the partisan prediction task. Examples include 'Abortion' and 'Contraception' as well as a cluster of issues including 'Ferguson Garner', 'Freddy Gray' and 'Trayvon'. Also, issues such as 'Common Core' seem to decrease performance across the board. This shows once again that each issue

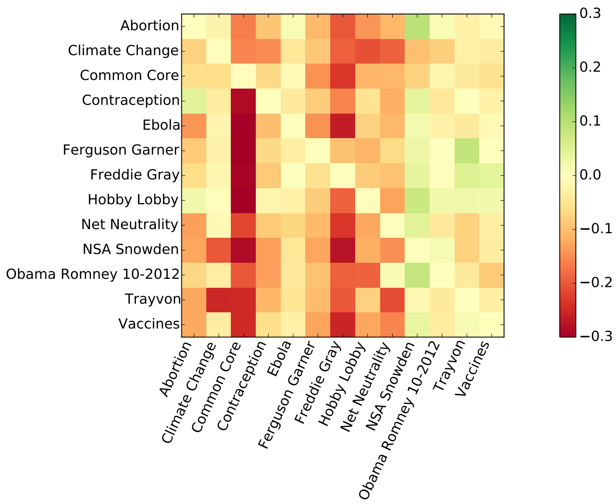



Figure 5: Relative changes in accuracy when using a different issue as training (horizontal axis) to predict a different issue (vertical axis).

is framed differently, with no obvious common denominator. We have experimented using the moral foundations to predict the partisan side from an issue and this did not consistently improve over baseline performance on all issues.

## 6. Conclusion

This paper presents a large scale analysis of framing in partisan news sources. We explored differences in word usage across sides and shown that these can be used to predict the partisan side for article on unknown bias. We uncovered meaningful differences in themes between the two sides, highlighting that either side prefers different words for the same event (e.g. 'riot' vs 'uprising') and that they are focused mostly on combating the other point of view. By analyzing a broad range of issues, our analysis showed that word usage patterns have limited general

We used the moral foundations theory in order to uncover a deeper psychological motivation behind news framing. Our analysis used a hand-crafted lexicon to quantify these. We have shown that the partisan sides differ in their expression of moral foundations, with liberals endorsing the 'care/harm' and 'fairness' foundations and conservatives endorsing the 'loyalty' and 'authority' foundations, as posited under this theory. However, the 'purity' foundation was not adequately captured by our analysis. Intriguingly, while liberals were concerned with both the vice and virtue aspects in their moral foundations, conservatives seemed to focus only on the vice aspect, denouncing the lack of loyalty and respect for authority.

We plan to continue studying the moral foundations theory using quantitative methods in our goal to show that there are deeper psychological aspects to political framing that are universal across issues. Other directions include analyzing the posting behavior of partisan Twitter users who took moral foundations assessments and relating the spatial distribution of morality relevant posts to voting behavior (both within and across parties). We also plan to further quantitatively testing and improve the moral foundation dictionaries using crowd-sourced annotations of morality.

## Acknowledgements

The authors acknowledge the support of the Templeton Religion Trust, grant TRT-0048. This research was conducted with the help of the Media Cloud team and especially Hal Roberts, who offered support with using the platform.

## 7. References

Al Zamal, F., Liu, W., and Ruths, D. (2012). Homophily and latent attribute inference: Inferring latent attributes of twitter users from neighbors. In *Proceedings of the 6th international AAAI Conference on Weblogs and Social Media*, ICWSM.

Baumer, E., Elovic, E., Qin, Y., Polletta, F., and Gay, G. (2015). Testing and comparing computational approaches for identifying the language of framing in political news. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*, NAACL.

Card, D., Boydstun, A., Gross, J., Resnik, P., and Smith, N. (2015). The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*, ACL.

Cohen, R. and Ruths, D. (2013). Classifying political orientation on twitter: It's not easy! In *Proceedings of the 7th international AAAI Conference on Weblogs and Social Media*, ICWSM.

Feinberg, M. and Willer, R. (2015). From gulf to bridge: When do moral arguments facilitate political influence? *Personality and Social Psychology Bulletin*.

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience.* Harvard University Press.

Graham, J., Haidt, J., and Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96(5):1029–1046.

Haidt, J. and Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *Social Justice Research*, 20(1):98–116.

Haidt, J. and Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, 133(4):55–66.

Haidt, J. (2000). The positive emotion of elevation. *Prevention & Treatment*, 3(1).

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological Review*, 108:814–834.

Iyer, R., Koleva, S., Graham, J., Ditto, P., and Haidt, J. (2012). Understanding libertarian morality: The psychological dispositions of self-identified libertarians. *PLoS ONE*, 7(8), 08.

Nguyen, V.-A., Boyd-Graber, J., Resnik, P., and Miler, K. (2015). Tea party in the house: A hierarchical ideal point topic model and its application to republican legislators in the 112th congress. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*, ACL.

Pennacchiotti, M. and Popescu, A.-M. (2011). Democrats, republicans and starbucks afficionados: User classification in twitter. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD, pages 430–438.

Rao, D., Yarowsky, D., Shreevats, A., and Gupta, M. (2010). Classifying latent user attributes in twitter. In *Proceedings of the 2nd International Workshop on Search and Mining User-generated Contents*, SMUC.

Rozin, P., Lowery, L., Imada, S., and Haidt, J. (1999). The cad triad hypothesis: A mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity). *Journal of Personality and Social Psychology*, 76(4).

Scheufele, D. A. (1999). Framing as a theory of media effects. *Journal of Communication*, 49(4):103–122.

Sylwester, K. and Purver, M. (2015). Twitter language use reflects psychological differences between democrats and republicans. *PLoS ONE*, 10(9), 09.

Tsur, O., Calacci, D., and Lazer, D. (2015). A frame of mind: Using statistical models for detection of framing and agenda setting campaigns. In *Proceedings of the 53rd*

*Annual Meeting of the Association for Computational Linguistics*, ACL.

Volkova, S., Coppersmith, G., and Van Durme, B. (2014). Inferring user political preferences from streaming communications. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, ACL, pages 186–196.

Volkova, S., Bachrach, Y., Armstrong, M., and Sharma, V. (2015). Inferring latent user properties from texts published in social media (demo). In *Proceedings of the Twenty-Ninth Conference on Artificial Intelligence*, AAAI.

Yano, T., Resnik, P., and Smith, N. A. (2010). Shedding (a thousand points of) light on biased language. In *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, NAACL, pages 152–158.