Moral Foundations of U.S. Political News Organizations

William E. Padfield[1], Erin M. Buchanan, Ph.D.[2], & Kayla N. Jordan, Ph.D.[2]

[1] Missouri State University

[2] Harrisburg University of Science and Technology

Author Note

William Padfield completed this project as part of his thesis requirements for a master's degree candidate in Psychology at Missouri State University. Submitted to Meta-Psychology. Participate in open peer review by commenting through hypothes.is directly on this preprint. The full editorial process of all articles under review at Meta-Psychology can be found following this link: https://tinyurl.com/mp-submissions

Correspondence concerning this article should be addressed to William E. Padfield, 901 S. National Ave, Springfield, MO, 65897. E-mail: wpadfield@gmail.com

Abstract

Partisan differences and diviseness have become an increasing hot topic in psychological research. Many theories have been proposed to explain these differences and divisions including Moral Foundations Theory. The current research seeks to use a linguistic measure of Moral Foundations, the Moral Foundations Dictionary (MFD), to test the theory in terms of predicted partisan differences. Through web scraping, we extracted articles from popular partisan news sources' websites, calculated MFD word frequencies, and identified words' respective valences. This process attempts to uncover news outlets' positive or negative endorsements of certain moral dimensions concomitant with a particular ideology. In Experiment 1, we gathered political articles from four sources. We were unable to reveal significant differences in moral endorsements, but we solidified the method to be employed in further research. In Experiment 2, we expanded their number of sources to 10 and analyzed articles that pertain to two specific topics: the 2018 confirmation hearings of U.S. Supreme Court Justice Brett Kavanaugh and the partial U.S. Government Shutdown of 2018-2019. Once again, no significant differences in moral endorsements were found. Together with past work, the results shed doubt on the validity of the MFD as a reliable measurement tool.

Keywords: politics, morality, psycholinguistics

Moral Foundations of U.S. Political News Organizations

The field of moral psychology has a long history with modern moral psychology beginning in the late 1960s with Lawrence Kohlberg's theory of moral development (Kohlberg & Hersh, 1977). Since then, Kohlberg's theory has been highly criticized with numerous theories and hypotheses proposed to replace it. One of these theories is Moral Foundations Theory (MFT) proposed by Jonathan Haidt and Jesse Graham (Haidt & Graham, 2007). While the theory itself has been widely criticized and debated, the goal of the current work is to tackle a related measurement issue, the Moral Foundations Dictionary (MFD) (Graham, Haidt, & Nosek, 2009). Graham et al. (2009) developed a lexicon-based linguistic measure of MFT arguing that moral concerns and the propositions of MFT could be measured and tested with the words people use. Here we first review evidence and criticisms against this assumption. We then provide our own test of the MFD using partisan news articles.

Moral Foundations Dictionary

Haidt and Jesse (2007) formulated MFT as a method by which to capture the entirety of humans' moral domain. We argued older theories of moral psychology were focused primarily on issues of justice, fairness, and caring - individually focused foundations of morality that align with the beliefs of political liberals. In other words, moral psychology ignored the valid moral foundations of conservatives. MFT holds that people's moral domain can be mapped by quantifying their endorsement of five moral foundations: *harm/care*, *fairness/reciprocity*, *ingroup/loyalty*, *authority/respect*, and *purity/sanctity* (Haidt & Graham, 2007).

In order to capture language's role in moral and political reasoning, Graham et al. (2009) formulated the MFD in order to capture moral reasoning and justification as used in speech and text. The MFD is composed of 259 words, with around 50 words assigned to each of the five foundations. Graham et al. (2009) created a preliminary list of words that

they believed would be associated with the five foundations. Then, using the Linguistic Inquiry and Word Count [LIWC; Pennebaker, Booth, and Frances (2007)] computer program, they analyzed transcripts of liberal and conservative Christian sermons in order to obtain frequencies of the occurrence of words from we' initial list.

Similar to previous research on MFT, liberal ministers used *harm*, *fairness*, and *ingroup* words more often than conservative ministers. Conversely, conservative ministers used *authority* and *purity* words more often than liberal ministers. However, conservative ministers did not use *ingroup/loyalty* words more than liberals. Rather, liberal ministers used words pertaining to *ingroup/loyalty*, but in contexts that promote rebellion and independence - causes *opposite* to positive endorsements of that foundation (Graham et al., 2009).

Critiques of Moral Foundations

MFT has received criticism on the grounds that its assumptions regarding moral intuitions have little empirical basis. Suhler, Churchland, and Joseph (2011) criticized the content and taxonomy of the five foundations and question whether or not the foundations are sufficiently distinct as to stand as their own foundation. Likewise, Gray and Keeney (2015) argues that the measurement of the moral foundations confounds morality with other constructs (such as weirdness) and the moral foundations lack statistical and conceptual distinction. Schein and Gray (2015) argues that harm-based morality is the most parsimonious and logical explanation for MFT findings.

Beyond the critiques of the theory itself, many have also provided strong critiques of the measurement using MFD.Conducting two close conceptual replication and six extension studies, Frimer (2020) found that the liberal-conservative differences found in the original study replicated in only 30% of cases and the effect sizes were over 30 times smaller than reported in Graham et al. (2009). In a meta-analysis, the theorized

differences were only found for authority and purity (e.g., conservatives used more authority and purity words). Loyalty was contrary to theoretical predictions (e.g., used more by liberals), and no differences were found for harm and fairness. Other studies have also found conflicting findings which fail to support the basic assertions of MFT. Exploring the political Twitterverse, Sterling and Jost (2018) found differences in the MFD depend on political sophistication. For example, harm words were used more by liberals only if they had low levels of political sophistication; at high levels of political sophistication, conservatives used more harm words. While the MFD has not been extensively used (at least in published research), studies which have use the MFD to test partisan differences have found mixed results in terms of supporting MFT (Clifford & Jerit, 2013; Frimer, Tell, & Haidt, 2015; Sagi & Dehghani, 2014).

The most successful attempts to validate a linguistic measure of moral foundations have used more complex methods (compared to the original dictionary-based approach). These attempts generally fall into two categories: human annotations and semantic representations. Hopp, Fisher, Cornell, Huskey, and Weber (2021) represents the first approach where instead of relying on 'experts,' a large sample of lay people were crowdsourced to manually annotate document for content relevant to each foundation. While resource-intensive, this method better captures how people may encounter and judge moral issues in everyday life leading to a more ecologically valid dictionary. Garten, Boghrati, Hoover, Johnson, and Dehghani (2016) represents the second categories which instead relies on data-driven semantic analysis. Using shallow neural net models like word2vec (Mikolov, Chen, Corrado, & Dean, 2013), moral foundations were measured using semantic similarity vectors rather than simple word counts.

As a further conceptual replication of Frimer (2020), we test the liberal-conservative difference proposed by MFT using partisan media content and the MFD. We first test differences in a general news corpus compared liberal sources (National Public Radio

[NPR] and the New York Times [NYT]) to conservative sources (Fox News and Breitbart). Second, we examined news about specific political events: Brett Kavaugh's U.S. Supreme Court nomination and the U.S. Government Shutdown in 2018-2019. In the second experiment, the news sources were expanded to cover NYT, NPR, Slate, Huffington Post, and Politico on the liberal side and Fox News, Breitbart, The Rush Limbaugh Show, The Blaze, and Sean Hannity on the conservative side. To address potential issues of measurement with the MFD, the MFD was combined with previous research by the current authors (see below) and weighted by valence to create weighted percentages to better specify endorsement. Like Frimer (2020), results were judged based on the direction of the liberal-conservative difference compared MFT predictions and the size of the effect compared to past studies.

Experiment 1

Method

For Experiment 1, we approached the study with the intention to answer a method question. That is, this portion of the current research was conducted in order to solidify the best method by which to analyze political news text under the MFT framework while also alleviating some of the aforementioned valence problem observed in the MFD. We hypothesized the news sources generally perceived as liberal leaning (*NPR* and *The New York Times*) would contain MFD words and valences indicating endorsements of the individualizing moral foundations (*harm/care* and *fairness/reciprocity*). Additionally, we hypothesized the two sources generally perceived to be conservative leaning (*Fox News* and *Breitbart*) would feature MFD words and valences indicating equal endorsement of all five foundations.

Sources

Political articles were collected from the websites of four notable U.S. news sources, a process known as web scraping. The sources were *NYT*, *NPR*, *Fox News*, and *Breitbart*. They were selected for their widespread recognition and the fact political partisans have strong preferences for some sources over others. We determined the political lean of each source by referencing Mitchell, Matsa, Gottfried, and Kiley (2014)'s article demonstrating the self-reported ideological consistency represented by the consumers of several news sources. In general, the *NYT* and *NPR* are preferred by consumers reporting a liberal bias or lean. In contrast, *Fox News* and *Breitbart* are believed to have a conservative bias or lean. Mitchell et al. (2014)'s article presented political ideology as a scale ranging from "consistently liberal" to "consistently conservative." In between these extremes lie more moderate positions, including "mostly liberal," "mixed," and "mostly conservative." Owing to the lower number of sources analyzed herein, we elected to categorize the sources as either "liberal" and "conservative" in order to form a basis for comparison.

Political articles in particular were identified and subsequently scraped by including the specific URL directing to each source's political content in the *R* script. For example, rather than scrape from nytimes.com, which would return undesired results (non-political features, reviews, etc.), we instead included nytimes.com/section/politics so that more or less exclusively political content was obtained. All code for this manuscript can be found at https://osf.io/5kpj7/, and the scripts are provided inline with this manuscript written with the *papaja* library (Aust & Barth, 2017).

Identification of the sources' political URLs presented a problem for two of the sources owing to complications with how their particular sites were structured. While in the multi-week process of scraping articles, we noticed word counts for *NPR* and *Fox News* were not growing at a similar pace as those from the *NYT* and *Breitbart.* Upon investigation, we found another, more robust URL for political content from NPR: their

politics content "archive." The page structure on NPR's website was such that only a limited selection of articles is displayed to the user at a given time. Scraping both the archive and the normal politics page ensured we were obtaining most (if not all) new articles as they were published. We later ran a process in order to exclude any duplicate articles. *Fox News* presented a similar issue. We discovered *Fox News* utilized six URLs in addition to the regular politics page. These URLs led to pages containing content pertaining the U.S. Executive Branch, Senate, House of Representatives, Judicial Branch, foreign policy, and elections. Once again, duplicates were subsequently eliminated from any analyses.

Materials

Using the *rvest* library in the statistical package *R*, we pulled body text for individual articles from each of the aforementioned sources (identified using CSS language) and compiled them into a dataset (Wickham, 2016). Using this dataset, we identified word count and average word count per source. This process was completed once daily starting in February 2018 until March 2018. Starting in mid-March 2018, the process was completed twice daily - once in the morning and again in the evening. Data collection was terminated once 250,000 words per source was collected in April 2018.

Data analysis

Once data collection ended, the text was scanned using the *ngram* package in *R* (Schmidt, Gonzalez-Cabrera, & Tomasello, 2017). This package includes a word count function, which was used to remove articles that came through as blank text, as well as to eliminate text picked up from the Disqus commenting system used by certain websites. At this point, duplicate articles were discarded.

The article text was processed using the *tm* and *ngram* packages in *R* in order to render the text in lowercase, remove punctuation, and fix spacing issues (Feinerer &

Hornik, 2017). The individual words were then reduced to their stems (i.e., *abused* was stemmed to *abus*). The same procedure was applied to the MFD words and the words in the Warriner, Kuperman, and Brysbaert (2013) dataset. Using the Warriner et al. (2013) dictionary, the words making up each of the five foundations in the MFD were matched to their respective valence value.

Concurrent research by Jordan, Buchanan, and Padfield (2019) assesses the validity of both the MFQ and the MFD through a multi-trait multi-method analysis of the two instruments using multiple samples. The instruments and foundation areas are being analyzed against one another, in order to test reliability, as well as against the Congressional Record in order to test predictive validity for political orientation. We were able to identify a number of potential new words that, if added to the MFD, could comprise a dictionary with greater validity, and less likelihood of zero percent texts, as this often occurs with the current MFD. Those results have informed this analysis, and their updated findings may change the underlying dictionary used in this analysis (albeit, we do not expect any changes in the results presented below).

The source article words were compiled into a dataset where they were matched up with their counterparts in the MFD along with their valence and a percentage of their occurrence. Therefore, for each article, the percentage of the number of *harm/care* words occurring in the articles were calculated, and this process was repeated for each of the foundations. Words' percent occurrence were multiplied by their z-scored valence. Valences were z-scored in order to eliminate any ambiguity regarding the direction of the valence. Positive values indicate positive valence, and negative values indicate negative valence. Words were categorized in accordance to their MFD affiliation, creating a weighted sum for each moral foundation.

Results

Descriptive Statistics

We calculated descriptive statistics for each news source in order to understand any and all fundamental linguistic differences in the sources' use of English. Statistics calculated included average z-scored valence of the unique words per article, number of articles per source, total number of words per source, average number of tokens (words) per article in each source, average number of types (unique words) per article in each source, and mean readability level per source. Readability statistics were calculated using the Flesch-Kincaid Grade Level Readability formula (Kincaid, Fishburne, Robert P., Richard L., & Brad S., 1975). Readability is calculated using a formula where the total number of syllables, words, and sentences in a given passage are determinants of its difficulty. The obtained value is intended to match up with the U.S. grade level at which one should be able to comfortably read the passage (Kincaid et al., 1975). For example, a text with a readability score of 11 should be easily read by a U.S. high school junior.

As seen in Table 1, the sources are similar in some aspects yet different in others. Valence appears to be slightly positive across all sources. The large standard deviations seem to indicate little to no presence of a difference in valence across sources. *NYT* published the greatest number of articles as well as total words. *Breitbart* featured the lowest number of articles, and *NPR* the lowest number of total words from all articles. Per individual article, however, *Breitbart* appears to feature the highest average number of words as well as unique words. Once again the standard deviations call into question any apparent differences between sources. Finally, *Fox News* articles had the lowest reading grade level on average, while *NYT* had the highest. This result might be attributable to the greater number of tokens in the average *NYT* article compared to *Fox News*. The standard deviations for readability suggest the presence of a diverse array of articles for each source, ranging from low to high reading level. Large standard deviations suggest the

sources feature a lot of overlap between them in their representation of scores.

Table 1

Experiment 1 - Descriptive Statistics by Source

Source	M_V	SD_V	$N_{Article}$	N_{Words}	M_T	SD_T	M_{Ty}	SD_{Ty}	M_{FK}	SD_{FK}
NPR	0.28	0.23	695	302977	435.94	642.63	191.96	192.28	14.00	3.93
New York Times	0.30	0.13	406	452579	1114.73	511.86	454.27	154.58	16.42	3.36
Breitbart	0.29	0.18	1437	722022	502.45	347.90	243.35	120.75	18.57	7.89
Fox News	0.29	0.17	503	296779	590.02	528.60	283.56	189.00	17.25	7.21

Note. Readability statistics were calculated using the Flesch-Kincaid Grade Level readability formula. V = Valence, T = Tokens or total words, Ty = Types or unique words, FK = Flesch-Kincaid

Inferential Statistics

To analyze if news sources adhered to differences in word use based on their target audience, we utilized a multilevel model (MLM) to analyze the data. MLM is a regression technique that allows one to control for the repeated measurement and nested structured of the data, which creates correlated error (Gelman, 2006). Using the *nlme* library in *R* (Pinheiro, Bates, Debroy, Sarkar, & Team, 2017), each foundation's weighted percentage was predicated here by the political lean of the news source, using the individual news sources as a random intercept to control for the structure of the data.

The multilevel model did not indicate the presence of any significant or practical effect of political lean for any of the five moral foundations. The strongest effect size was observed for the *authority/respect* foundation, but the effect was in the opposite direction from what was originally hypothesized - liberal sources tended to use more *authority/respect* words than did conservative sources. Descriptive and test statistics, *p*-values and effect sizes (Cohen's *d*) can be found in Table 2. To interpret the weighted

scores, one can examine the mean and standard deviations for each. A zero score for the mean, with a non-zero standard deviation, would indicate a perfect balance of positive and negative words in each category, likely representing a neutral tone when all words are considered. Negative percentages would indicate more representation of the negative words in the MFD area, while positive percentages indicate an endorsement of the positive words in a MFD. Therefore, we suggest using the sign of the mean score to determine the directionality of the endorsement for the MFD (positive, neutral, negative), and the standard deviation to ensure that a zero score is not zero endorsement (i.e., a SD of zero indicates no words were used). Based on the weighted percent values for the five foundations, we observed that MFD words seem to make up a small portion of the article text. Furthermore, the observed percentages and means appear to indicate a generally positive endorsement of all five foundations across both liberal and conservative sources.

Table 2

Experiment 1 Results - Multilevel Model

Foundation	M_C	SD_C	M_L	SD_L	t	p	d
Harm/Care	0.50	2.21	0.49	2.21	-0.21	.851	0.01
Fairness/Reciprocity	1.13	1.38	1.11	1.38	-0.42	.713	0.02
Ingroup/Loyalty	1.28	1.63	1.34	1.63	0.30	.793	-0.04
Authority/Respect	0.72	1.62	1.06	1.62	3.16	.087	-0.20
Purity/Sanctity	1.11	1.48	1.27	1.48	2.38	.140	-0.09

Note. For mean and standard deviation values, 'C' and 'L' refer to 'conservative' and 'liberal,' respectively

Discussion

The results obtained in Experiment 1 fail to support the differences predicted by MFT. First, differences between liberal and conservative news sources failed to reach statistical or practical (effect size) significance for any foundations. Second, looking purely

at the direction of the differences, the results run completely contrary to the differences expected by MFT. Conservative sources scored higher on concern for harm and fairness while liberal sources scored higher on concern for loyalty, authority, and purity. Overall, the results lend strong support to the case against the MFD and weakens the case generally for MFT.

However, one possible limitation of the current data is the generalness of the corpus. The selection of the broad and amorphous topic of "political news" may have led to the scraping of large numbers of articles with little to no moral-centric content. To address this limitation, two changes that were subsequently employed in Experiment 2. First, we elected to include more news sources for web scraping and analysis in addition to the four used in Experiment 1. Second, we chose to focus data collection efforts on two heavily moralized events in the Trump administration: (1) the nomination and confirmation of Justice Brett Kavanaugh to the U.S. Supreme Court and the U.S. government shutdown in December 2018 through January 2019 over disagreements about funding a U.S.-Mexico border wall. Hence, Experiment 2 is the best test of the MFT and the MFD in the context of partisan news.

Experiment 2

Method

In contrast to Experiment 1, we approached Experiment 2 with the intention to confirm the method employed was valid for the analysis of the scraped text as well as for any inferences drawn from the analyses. For Experiment 2, we hypothesized that news sources perceived as liberal will exhibit positive endorsements of the individualizing moral foundations (*harm/care* and *fairness/reciprocity*) in their articles reporting on both the Kavanaugh confirmation hearing as well as the 2018-2019 government shutdown. News sources perceived as conservative are hypothesized to positively endorse all five foundations equally in their coverage of the Kavanaugh hearing and the government shutdown. We

tested the hypothesis by analyzing the content scraped from news sources' web pages spanning the two weeks before (September 13, 2018) and two weeks after (October 11, 2018) Kavanaugh's confirmation hearing, owing to its prominence in the news. Likewise, we analyzed content spanning two weeks before the start of the government shutdown (December 8, 2018) to two weeks following the end of the shutdown (February 8, 2019). The content will be analyzed for valence and moral alignment under MFT.

Sources

Articles pertaining to the Brett Kavanaugh Supreme Court confirmation hearing and the 2018-2019 U.S. Government shutdown were scraped from the websites of 10 U.S. news sources. As in Experiment 1, these sources were selected owing to their favorability among political partisans according to Mitchell et al. (2014). The sources favored by the highest proportion of consistent liberals were the *NYT*, *NPR*, *Slate*, *Huffington Post*, and *Politico* (Mitchell et al., 2014). The sources favored by the highest proportion of consistent conservatives included *Fox News*, *Breitbart*, *The Rush Limbaugh Show*, *The Blaze*, and *Sean Hannity*. Political articles referencing Brett Kavanaugh's nomination process were identified and subsequently scraped by including the URL for each source's coverage of the nomination in the *R* script.

Materials

Using the *rvest* and *RSelenium* libraries, we pulled body text for individual articles from each of the aforementioned 10 news sources and compiled them together (Harrison & Kim, 2020; Wickham, 2016). Using this dataset, we identified word count and average word count per source. This process was run for articles pertaining to Kavanaugh's nomination that were published between September 13, 2018 and October 11, 2018 inclusive. This date range was selected in reference to the widely-publicized and viewed nomination hearing on September 27, 2018. We set the start date at September 13 (two weeks before the hearing) and the end date at October 11 (two weeks after the hearing) so

that we could capture a large amount of data (roughly one month) during which Kavanaugh's nomination was at its peak saturation in news coverage.

The same process was followed for scraping articles related to the partial U.S. Government shutdown of 2018-2019. The articles scraped were published between December 8, 2018 and February 8, 2019 inclusive. Once again, we elected to scrape articles published two weeks before and after the event in question in order to capitalize on the shutdown's saturation in American news media.

Data analysis

As in Experiment 1, the text was scanned with *ngram* (Schmidt et al., 2017). Again, blank articles, text from the Disqus system, and duplicate articles were removed. The text was processed and stemmed in order to convert to a usable form for further analysis (Feinerer & Hornik, 2017). Each individual word was reduced to its stem (i.e., *diseased* was stemmed to *diseas*). Once again, the same procedure was applied to the MFD words and the words in the Warriner et al. (2013) dataset. We obtained the words' percent occurrence in the text. Percents were multiplied by *z*-scored valence and categorized into their proper MFD category.

<div align="center">

Results

</div>

Descriptive Statistics

We calculated descriptive statistics for each news source per topic in order to reveal the presence (if any) of linguistic differences in the sources' use of language. As in Experiment 1, statistics calculated include *z*-scored valence, number of articles per source, total words per source, mean tokens per article in each source, mean types per article in each source, and mean readability level (using the Flesch-Kincaid Grade Level Readability formula) per source (Kincaid et al., 1975).

Table 3 displays the descriptive statistics for sources' writing on the Kavanaugh confirmation hearing. The sources were similar in most basic linguistic aspects, except for number of articles. For example, *Sean Hannity* appears to have published only 27 articles while *Breitbart* published 757 articles on this topic. Valence was found to be slightly positive across all sources. *Fox News* produced the most total words with the most tokens on average. This is likely due to the fact *Fox News* transcribes many of their videos and publishes them in article form. *Politico* featured the highest number of types on average. *Rush Limbaugh* featured the lowest readability score on average by grade level while *Slate* featured the highest grade-level readability score. The large standard deviations for these statistics, however, preclude conclusions regarding differences in the sources' use of language, as there is likely a lot of overlap between sources' use of language.

Table 3

Kavanaugh - Descriptive Statistics by Source

Source	M_V	SD_V	$N_{Article}$	N_{Words}	M_T	SD_T	M_{Ty}	SD_{Ty}	M_{FK}	SD_{FK}
Huffington Post	0.27	0.12	552	359046	650.45	462.21	283.07	129.23	10.68	1.89
NPR	0.22	0.25	366	108605	296.73	499.10	128.12	172.51	12.55	3.67
New York Times	0.33	0.13	653	723569	1108.07	570.31	461.26	174.20	9.52	1.79
Politico	0.32	0.10	689	1069292	1551.95	1045.47	614.66	358.68	12.10	2.66
Slate	0.29	0.12	272	229896	845.21	612.22	332.42	168.17	12.30	2.48
The Blaze	0.25	0.13	277	128097	462.44	149.66	210.52	53.14	10.73	1.85
Breitbart	0.30	0.13	757	375848	496.50	609.91	230.03	153.62	10.87	2.12
Fox News	0.29	0.11	646	1304048	2018.65	2709.77	534.47	404.21	9.71	1.86
Sean Hannity	0.31	0.11	27	5926	219.48	145.98	121.04	57.26	11.76	2.40
Rush Limbaugh	0.38	0.13	172	267067	1552.72	1258.51	419.00	225.80	9.17	8.86

Note. Readability statistics were calculated using the Flesch-Kincaid Grade Level readability formula. V = Valence, T = Tokens or total words, Ty = Types or unique words, FK = Flesch-Kincaid

Table 4 displays descriptive statistics for articles about the partial government shutdown. Like the Kavanaugh hearing, the sources were similar in average valence (slightly positive). Once again, there was variation in the number of articles published by each source on this topic. *Sean Hannity*, *Rush Limbaugh*, and *The Blaze* published fewer than 100 articles while *Fox News* published 1,013 articles. *Fox News* again featured the most total words and mean tokens, but this is likely due to to the presence of a large amount of video transcriptions that the organization published as articles. *Politico* had the most types on average. For this topic, *Fox News* featured the lowest reading grade level while *Slate* featured the highest reading grade level. For each statistic, the excessively high standard deviations render any assertions regarding linguistic differences inconclusive on a descriptive level due to the aforementioned overlap in sources' language use.

Table 4

Government Shutdown - Descriptive Statistics by Source

Source	M_V	SD_V	$N_{Article}$	N_{Words}	M_T	SD_T	M_{Ty}	SD_{Ty}	M_{FK}	SD_{FK}
Huffington Post	0.30	0.13	432	242220	560.69	425.52	258.27	130.54	10.68	1.85
NPR	0.23	0.22	434	151114	348.19	457.34	155.53	171.30	11.20	3.16
New York Times	0.33	0.13	752	836811	1112.78	532.72	472.95	177.57	10.10	1.74
Politico	0.29	0.10	222	349499	1574.32	875.63	638.32	313.34	11.31	1.26
Slate	0.30	0.12	117	85254	728.67	464.34	301.71	138.74	11.87	2.53
The Blaze	0.26	0.13	98	38983	397.79	102.10	185.66	41.93	10.70	1.85
Breitbart	0.34	0.15	309	102886	332.96	238.70	168.31	81.59	10.71	2.05
Fox News	0.35	0.13	1013	2799217	2763.29	3405.44	637.30	494.00	9.43	1.95
Sean Hannity	0.29	0.14	63	10311	163.67	33.59	95.13	16.59	13.36	4.81
Rush Limbaugh	0.37	0.12	78	152630	1956.79	1544.14	482.79	252.79	9.77	7.64

Note. Readability statistics were calculated using the Flesch-Kincaid Grade Level readability formula. V = Valence, T = Tokens or total words, Ty = Types or unique words, FK = Flesch-Kincaid

Inferential Statistics

We utilized a multilevel model (MLM) to analyze if news sources leveraged different vocabularies based on target audience in the same manner as Experiment 1. Each foundation's weighted percentage was predicted by the source's political lean, using the individual source as a random intercept to control for the nested data structure. Two separate MLMs were constructed from datasets compiled for each topic of interest: the Kavanaugh hearing and the partial government shutdown of 2018-2019.

For the Kavanaugh topic, the multilevel model indicated the presence of a statistically significant effect for *harm/care*, but the practical effect denoted by Cohen's *d* was found to be small. There were no other significant or practical effects of political lean for any of the other four moral foundations. The effect for *harm/care* was in the hypothesized direction with liberal sources tending to use more positively *harm/care* words than conservative sources. Descriptive and test statistics, *p*-values and effect sizes (Cohen's *d*) can be found in Table 5.

Table 5

Kavanaugh Results - Multilevel Model

Foundation	M_C	SD_C	M_L	SD_L	t	p	d
Harm/Care	0.39	1.35	0.66	1.35	2.89	.020	-0.19
Fairness/Reciprocity	1.26	1.27	1.06	1.27	-1.58	.154	0.16
Ingroup/Loyalty	1.15	1.22	1.06	1.22	-0.63	.549	0.07
Authority/Respect	0.85	0.99	0.75	0.99	-0.93	.380	0.10
Purity/Sanctity	0.90	1.28	1.06	1.28	0.88	.406	-0.11

Note. For mean and standard deviation values, 'C' and 'L' refer to 'conservative' and 'liberal,' respectively

For news articles about the partial U.S. Federal Government Shutdown of 2018-2019, there were no significant or practical effects of political lean for the moral foundations (see

Table 6). A small-to-medium effect size was observed for *authority/respect.* The effect was in the predicted direction as well, as conservative sources tended to offer more positive endorsements of the foundation than liberal sources. This effect was similar to Experiment 1 in which the largest effect size was observed for *authority/respect.* As noted before, the effect found in Experiment 1 was in the opposite direction as what was hypothesized. Thus, it is difficult to draw comparisons between the two studies despite the similar pattern for effect size. Owing to a lack of similar effects for either Experiment 1 or the Kavanaugh topic, there is doubt as to whether or not a practical or generalized effect exists for *authority/respect.*

Based on the weighted percent values for the five foundations applied to both topics, MFD words seem to make up little of the article text. A similar pattern was observed for the results in Experiment 1. As in Experiment 1, the percentages and means seem to indicate a generally positive endorsement of all five moral foundations across both political leanings.

Table 6

Government Shutdown Results - Multilevel Model

Foundation	M_C	SD_C	M_L	SD_L	t	p	d
Harm/Care	0.91	1.27	1.01	1.27	1.68	.132	-0.07
Fairness/Reciprocity	1.22	1.05	0.91	1.05	-0.76	.467	0.27
Ingroup/Loyalty	1.38	1.16	1.11	1.16	-0.48	.643	0.21
Authority/Respect	0.87	1.24	0.36	1.24	-1.38	.204	0.37
Purity/Sanctity	1.29	1.13	0.99	1.13	0.03	.979	0.22

Note. For mean and standard deviation values, 'C' and 'L' refer to 'conservative' and 'liberal,' respectively

Discussion

The results obtained in Experiment 2 are slightly more in line with expectations from MFT, however, in all but one case, the differences were not statistically significant. The only statistically significant difference was in the harm foundation; liberals did endorse harm more than conservatives during the Kavanaugh hearings, but the effect was smaller than expected given Graham et al. (2009). Generally for both the Kavanaugh hearings and the Government shutdown, the differences were in the expected direction (liberals endorsing harm more and conservatives endorsing authority and loyalty more), but the differences were not statistically significant and the practical effects much smaller than reported in Graham et al. (2009). While not significant, fairness was endorsed more by conservatives in both cases. In contrast, purity was endorsed more by conservatives for the government shutdown, but more by liberals for the Kavanaugh hearing though neither difference was statistically significant.

Together with Experiment 1 and past work, Experiment 2 further calls into question the usefulness and efficacy of the Moral Foundations as well as cast doubt on the predictions made by MFT. One issue that Experiment 2 brings up is that moral differences between liberals and conservatives may be moderated by a number of factors as shown in Sterling and Jost (2018). For example, here we have liberals concerned with purity for the Kavanaugh hearings (a moral bad for liberals and a moral good for conservatives) and conservatives concerned with purity during the government shutdown (a moral bad for conservatives and a moral good for liberals). The original predictions made by the Moral Foundations are very simplistic and likely much more complex in reality.

Conclusions

Within the theoretical framework of MFT (Haidt & Graham, 2007), we attempted to devise a method leveraging the MFD (Graham et al., 2009) in order to quantify political bias stemming from content published by several prominent American news sources. In

Experiment 1,the results obtained were not significant in any statistical or practical sense. There was a small effect found for the *authority/respect* foundation according to Cohen's *d*, but the effect suggested endorsement for that foundation opposite to what was hypothesized. In Experiment 2, the direction of the differences were generally in line with MFT, however, the difference were very small and lacks statistically significance. Building on Frimer (2020) and other critiques of MFT and the MFD, we show that MFD is not a useful tool for measuring moral language while also calling into question the validity of MFT in terms of theorized partisan differences.

Despite the fact the results regarding political bias were inconclusive, we still retain confidence in the overall structure of the methodology established in the current study. Specifically, the procedure for scraping text from the web, processing, stemming, and weighting the scores with valence seems to represent a solid method for preparing a high quantity of text passages for data analysis. We implemented valence as an indicator of the directionality of endorsement due to the inherent ambiguity of simply calculating MFD word percent occurrence. This method served both to augment the face validity of the MFD by incorporating valence (thus reducing ambiguity) and to generate a score that is easy to understand and analyze.

However, political lean and other important political/moral constructs may be communicated through methods other than through the use of particular words. As mentioned before, MFD is a measure developed from MFT (Graham et al., 2009). Based upon the results obtained, it might be necessary to investigate alternative instruments that could better elucidate the differences of interest. Going beyond specific instruments, other theoretical perspectives may be more equipped to explain political differences in discourse. Likewise, an atheoretical approach in which large quantities of data are collected from which theories are formulated may be best suited to this area of research.

In critiquing MFT, Schein and Gray (2015) proposed an alternative theory explaining

moral choices and potential partisan differences: Dyadic Theory of Morality. Across many studies, researchers find that supposed differences in moral foundation can be more easily explained by a harm-focused morality where someone/thing is harmed by someone/thing else Schein & Gray (2015). Rather than liberals and conservatives having difference conceptions of morality, dyadic morality argues that they simply identify harm differently. For example, in the case of abortion, liberals tend to identify the primary harm to the mother leading to a pro-choice position whereas conservative tend to identify the primary harm to the fetus leading to a pro-life position. While as of yet, no linguistic measurement of dyadic morality exists; such an approach may be able to better identify the systematic differences between liberals and conservatives.

Turning to atheoretical approaches, future studies aiming to uncover political lean in discourse may benefit from a bottom-up, data-driven approach. Researchers might derive substantive insights into political lean through gathering data after which they may formulate theories that explain systematic observations obtained from that data. The availability of text corpora along with methods for extracting large amounts of text from the internet (as was demonstrated in the current study) potentially make this a feasible option. Likewise, there are several approaches to analyzing such data, including linear models (like multilevel models) and network-style models such as latent semantic analysis (Landauer, 1998). Owing to the wealth of representative data as well as the sophistication of current analytic tools, there is high potential for the explanatory power of new theories involving political discourse.

Never before has political discourse represented such fertile ground for psychological research. The plethora of options for news sources has created not only an abundance of choice but also vast quantities of text data. Along with this recent increase in the amount of text information, there is now an obligation on the part of researchers to devise proper methods for analyzing that text. Solid methodologies must be constructed and periodically

improved to keep pace with evolving technologies. Likewise, a strong theoretical foundation is paramount to making sense of the current and future media ecosystem. Therefore, it is incumbent upon social scientists to continue investigating the information consumed by millions of Americans every day so that insights into the nature and consequences of political discourse can be more completely understood.

References

Aust, F., & Barth, M. (2017). papaja: Create APA manuscripts with R Markdown. Retrieved from https://github.com/crsh/papaja

Clifford, S., & Jerit, J. (2013). How words do the work of politics: Moral foundations theory and the debate over stem cell research. *The Journal of Politics*, *75*(3), 659–671. https://doi.org/10.1017/S0022381613000492

Feinerer, I., & Hornik, K. (2017). Text mining package. Retrieved from http://tm.r-forge.r-project.org/

Frimer, J. A. (2020). Do liberals and conservatives use different moral languages? Two replications and six extensions of Graham, Haidt, and Nosek's (2009) moral text analysis. *Journal of Research in Personality*, *84*, 103906. https://doi.org/10.1016/j.jrp.2019.103906

Frimer, J. A., Tell, C. E., & Haidt, J. (2015). Liberals condemn sacrilege too: The harmless desecration of Cerro Torre. *Social Psychological and Personality Science*, *6*(8), 878–886. https://doi.org/10.1177/1948550615597974

Garten, J., Boghrati, R., Hoover, J., Johnson, K. M., & Dehghani, M. (2016). Morality between the lines: Detecting moral sentiment in text. In *Proceedings of IJCAI 2016 workshop on Computational Modeling of Attitudes*.

Gelman, A. (2006). Multilevel (hierarchical) modeling: What it can and cannot do. *Technometrics*, *48*(3), 432–435. https://doi.org/10.1198/004017005000000661

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Gray, K., & Keeney, J. E. (2015). Disconfirming moral foundations theory on its own terms: Reply to Graham (2015). *Social Psychological and Personality Science*, *6*(8), 874–877. https://doi.org/10.1177/1948550615592243

Gray, K., Schein, C., & Ward, A. F. (2014). The myth of harmless wrongs in moral

cognition: Automatic dyadic completion from sin to suffering. *Journal of Experimental Psychology: General*, *143*(4), 1600–1615. https://doi.org/10.1037/a0036149

Haidt, J., & Graham, J. (2007). When morality opposes justice: Conservatives have moral intuitions that Liberals may not recognize. *Social Justice Research*, *20*(1), 98–116. https://doi.org/10.1007/s11211-007-0034-z

Harrison, J., & Kim, J. Y. (2020). RSelenium: R Bindings for 'Selenium WebDriver'. Retrieved from https://CRAN.R-project.org/package=RSelenium

Hopp, F. R., Fisher, J. T., Cornell, D., Huskey, R., & Weber, R. (2021). The extended moral foundations dictionary (eMFD): Development and applications of a crowd-sourced approach to extracting moral intuitions from text. *Behavior Research Methods*, *53*(1), 232–246. https://doi.org/10.3758/s13428-020-01433-0

Jordan, K. N., Buchanan, E. M., & Padfield, W. E. (2019). *A validation of the Moral Foundations Questionnaire and Dictionary*. Retrieved from https://osf.io/kt9yf/

Kincaid, J. P., Fishburne, Jr., Robert P., R., Richard L., C., & Brad S. (1975). Derivation of new readability formulas (automated readability index, fog count and Flesch reading ease formula) for Navy enlisted personnel. https://doi.org/10.21236/ADA006655

Kohlberg, L., & Hersh, R. H. (1977). Moral development: A review of the theory. *Theory Into Practice*, *16*(2), 53–59. https://doi.org/10.1080/00405847709542675

Landauer, T. K. (1998). Learning and Representing Verbal Meaning. *Current Directions in Psychological Science*, *7*(5), 161–164. https://doi.org/10.1111/1467-8721.ep10836862

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *arXiv:1301.3781 [Cs]*. Retrieved from http://arxiv.org/abs/1301.3781

Mitchell, A., Matsa, K. E., Gottfried, J., & Kiley, J. (2014). Political polarization & media habits | Pew Research Center. Retrieved from http://www.journalism.org/2014/10/21/political-polarization-media-habits/

Pennebaker, J. W., Booth, R. J., & Frances, M. E. (2007). Liwc2007: Linguistic inquiry

and word count. Austin, TX.

Pinheiro, J., Bates, D., Debroy, S., Sarkar, D., & Team, R. C. (2017). nlme: Linear and nonlinear mixed effects models. Retrieved from https://cran.r-project.org/package=nlme

Sagi, E., & Dehghani, M. (2014). Measuring moral rhetoric in text. *Social Science Computer Review*, *32*(2), 132–144. https://doi.org/10.1177/0894439313506837

Schein, C., & Gray, K. (2015). The unifying moral dyad: Liberals and conservatives share the same harm-based moral template. *Personality and Social Psychology Bulletin*, *41*(8), 1147–1163. https://doi.org/10.1177/0146167215591501

Schmidt, M. F. H., Gonzalez-Cabrera, I., & Tomasello, M. (2017). Children's developing metaethical judgments. *Journal of Experimental Child Psychology*, *164*, 163–177. https://doi.org/10.1016/j.jecp.2017.07.008

Sterling, J., & Jost, J. T. (2018). Moral discourse in the twitterverse: Effects of ideology and political sophistication on language use among US citizens and members of congress. *Journal of Language and Politics*, *17*(2), 195–221. https://doi.org/10.1075/jlp.17034.ste

Suhler, C. L., Churchland, P., & Joseph, C. (2011). Can innate , modular "foundations" explain morality? Challenges for Haidt's Moral Foundations Theory. *Journal of Cognitive Neuroscience*, *23*(9), 2103–2116. https://doi.org/10.1162/jocn.2011.21637

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, *45*(4), 1191–1207. https://doi.org/10.3758/s13428-012-0314-x

Wickham, H. (2016). Package 'rvest'. Retrieved from https://cran.r-project.org/package=rvest