

Portland State University

**PDXScholar**

---

Dissertations and Theses

Dissertations and Theses

---

1-12-2022

# Measuring Morality in the 2020 US Presidential Election

Scott Edward Atkins  
*Portland State University*

Follow this and additional works at: [https://pdxscholar.library.pdx.edu/open\\_access\\_etds](https://pdxscholar.library.pdx.edu/open_access_etds)



Part of the [Communication Commons](#), [Political Science Commons](#), and the [Psychology Commons](#)

**Let us know how access to this document benefits you.**

---

## Recommended Citation

Atkins, Scott Edward, "Measuring Morality in the 2020 US Presidential Election" (2022). *Dissertations and Theses*. Paper 5893.

<https://doi.org/10.15760/etd.7764>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: [pdxscholar@pdx.edu](mailto:pdxscholar@pdx.edu).

Measuring Morality in the 2020 US Presidential Election

by

Scott Edward Atkins

A thesis submitted in partial fulfillment of the  
requirements for the degree of

Master of Science  
in  
Communication

Thesis Committee:  
L. David Ritchie, Chair  
Cynthia-Lou Coleman  
Jeffrey Robinson

Portland State University  
2021

© 2021 Scott Edward Atkins

## **Abstract**

This study explores the moral content evident in speeches by 2020 US Presidential Candidates Joe Biden and Donald Trump. Drawing on Moral Foundations Theory (Haidt, 2013), I test the hypotheses that each candidate's moral content, as measured by the use of certain morally salient keywords, will fall along patterns based on their political affiliation. In testing these hypotheses, I also present a comparison of keyword analysis methods. The first uses a simple word count procedure alongside the Moral Foundations Dictionary 2.0, developed by Frimer et al. (2017), which scores a document based on the presence of words from each of the moral foundations. This method is a direct offshoot of Moral Foundations Theory, with earlier iterations having been used in development of the theory (Graham et al., 2009). The second method uses the extended Moral Foundations Dictionary developed by Hopp et al. (2020), which scores documents based on the moral scores of words derived from a crowd-sourced development project.

Results indicate some departures from the predicted model, but more striking is the finding that there appears to be little significant difference between the two candidates' overall pattern of moral keyword use. However, this pattern is not consistent across both methods of analysis. Thus, I also present a comparison of these methods and comment on underlying differences in operationalization that call into question whether they are truly measuring the same thing.

### **Dedication**

To my brother and sister, with whom I probably agree more than we all realize, and in memory of our mother, Linda Atkins, who certainly had opinions about morality and politics; 1946-2020.

## **Acknowledgments**

I would like to thank my wife, Katie, without whom my journey through higher education would have been much more challenging. Her support and sacrifices helped me in countless ways. Likewise, I would like to thank our two daughters for their ever-present source of energy and joy.

I am indebted to the members of my committee for their input that pushed this thesis towards something I can be proud of. I appreciate Dr. Coleman and Dr. Robinson's direction that kept the work grounded in theory and pointed towards furthering the understanding and application of new knowledge. Dr. Ritchie's persistent belief in the ability of his students has consistently given me a high standard to strive towards, even as I surprise myself at meeting that standard.

Finally, I would like to acknowledge my fellow graduate students in the Department of Communication. When we began our program in 2019, we had no idea how COVID-19 would soon change everything. I could not have come this far without the camaraderie and support we all shared, at first in person and eventually over Discord and Zoom. Thanks especially to Luke, Siri, Jamie, Emily, Colby, and Christian.

## Table of Contents

Abstract .....	i
Dedication .....	ii
Acknowledgments .....	iii
List of Tables .....	v
List of Figures .....	vi
Introduction .....	1
Literature Review .....	5
Experiment 1 .....	28
Experiment 2 .....	36
Experiment 3 .....	43
General Discussion .....	49
Limitations and Conclusions .....	53
References .....	55

## List of Tables

Table 1	
Descriptive Statistics of LIWC2015 Analysis Using the Moral Foundations Dictionary 2.0.....	32
Table 2	
Means, Standard Deviations, and One-Way Analysis of Variance of Moral Content Using LIWC2015 and the Moral Foundations Dictionary 2.0. ....	32
Table 3	
Within-Subject Comparisons Using LIWC2015 and the Moral Foundations Dictionary 2.0.....	34
Table 4	
Descriptive Statistics of LIWC2015 Analysis Using the Moral Foundations Dictionary 2.0 with Segmented Corpus. ....	37
Table 5	
Means, Standard Deviations, and One-Way Analysis of Variance of Moral Content Using LIWC2015 and the Moral Foundations Dictionary 2.0 with Segmented Corpus...38	
Table 6	
Descriptive Statistics of eMFDScore Analysis Using the Moral Foundations Dictionary 2.0 with Segmented Corpus. ....	41
Table 7	
Means, Standard Deviations, and One-Way Analysis of Variance of Moral Content Using eMFDScore and the Moral Foundations Dictionary 2.0 with Segmented Corpus..41	
Table 8	
Descriptive Statistics of eMFDScore Analysis Using the extended Moral Foundations Dictionary with Segmented Corpus. ....	45
Table 9	
Means, Standard Deviations, and One-Way Analysis of Variance of Moral Content Using eMFDScore and the extended Moral Foundations Dictionary with Segmented Corpus. ....	45
Table 10	
Within-Subject Comparisons Using eMFDScore and the extended Moral Foundations Dictionary. ....	47



**List of Figures**

Figure 1	
Boxplot Depicting Moral Keyword Analysis Using LIWC2015 and the Moral Foundations Dictionary 2.0.....	31
Figure 2	
Boxplot Depicting Moral Keyword Analysis Using eMFDScore and the Moral Foundations Dictionary 2.0.....	40
Figure 3	
Boxplot Depicting Moral Keyword Analysis Using eMFDScore and the extended Moral Foundations Dictionary.....	44

## Introduction

When Joseph Biden announced his campaign for the Democratic presidential nomination, he claimed that the 2020 election represented “a battle for the soul of the nation” (Joe Biden, 2019). This became his campaign slogan, a veritable battle cry to rally voters against sitting president Donald Trump, whom Biden claimed had “long ago forfeited any moral leadership in this country” (Biden, 2020). Despite the possibly confrontational tone inherent in a message about battle, Biden’s campaign emphasized the importance of unity among all Americans regardless of differences in political ideology. However, Biden only won the election with approximately 51% of all votes cast (“Fact check”, 2020). It appears that his appeal for unity was not enough to win over voters who may have already decided to vote for other candidates.

The explanation for Biden’s narrow victory may be as simple as voters forming their opinions and casting their ballots along party lines despite the potential merits of an opposing candidate. But what other factors may have been involved? Research linking politics and moral psychology suggests that liberals and conservatives often use different language to discuss the same topics, thus appealing to different moral foundations to make their cases (e.g., Feinberg & Willer, 2015; Haidt, 2013; Lakoff, 1996). Perhaps Biden’s messages were couched in inherently “liberal” or “democratic” language that dissuaded conservative voters? In the current study, I explore this possibility by using two methods of moral content analysis to discover whether Biden’s and Trump’s language contains partisan based moral content in patterns predicted by Moral Foundations Theory. Results indicate more apparent similarities than differences between the candidates, although the two measures do not agree on the overall patterns of moral

language, likely as a result of distinct differences in the underlying definition and measurement of moral content.

It seems plausible that Biden's criticism of Donald Trump's "moral forfeiture" was based, at least in part, on Trump's divisive reputation. During his presidency, Trump established a penchant for incivility directed towards political opponents, both public and private figures alike (Baker, 2017; Stohr, 2017). For example, Trump denounced players in the National Football League who knelt during the national anthem in protest of racial inequality, claiming that these players were unpatriotic and should be banned from the league (Belson & Davis, 2017). He derided his own staff for making choices that did not advance his interests, such as his Attorney General Jeff Sessions, who recused himself from an investigation into links between Trump and the Russian government (Gstalter, 2019). Trump mocked the parents of a fallen US soldier who spoke against him during the 2016 Democratic National Convention ("Fury...", 2016). He once claimed that his fame and fortune allowed him to sexually assault women (Taylor, 2016), a comment that he waved aside as "locker room talk" during the 2016 debates with candidate Hillary Clinton (Diaz, 2016). These are but a few examples of Trump's style of rhetoric, described by one media reporter as strategy based on the idea that there are "so many people to attack, so little time" (Wehner, quoted in Baker, 2017).

Trump's contemptuous rhetoric became a regular pattern, enabling similar attitudes and behaviors in his supporters (Stohr, 2017). During the Trump presidency, latent right-wing extremism surfaced, emboldened by a president who fails to condemn those who support his own causes yet threaten our nation's democratic institutions (Rakich, 2021). The most troubling example of this occurred on January 6, 2021. While

Congress was convened to officially count the electoral votes and certify the results of the 2020 election, President Trump led a rally in which he outlined the ways he felt the election had been “stolen” from him. He encouraged his supporters to march to the US Capitol in demonstration of support for those members of Congress that might uphold a Trump victory (Naylor, 2021). A large group of supporters did march, and in fact laid siege to the US Capitol Building, causing congress to evacuate or shelter behind locked doors for their safety.

Of course, political strife during the years of Trump’s presidency was not the sole fault of his supporters. Left-wing activists, notably the group ANTIFA (i.e. Anti-Fascist), exhibited similar displays of militancy, often in direct physical conflict with right-wing activists. Racial tensions were high throughout 2020, when the death of George Floyd, an unarmed Black man, at the hands of police officers sparked months-long demonstrations worldwide. Many of these demonstrations turned violent, as some left-wing agitators used the protests as opportunities to protest against police officers and promote some form of anarchistic ideology.

Research suggests that we are living in an increasingly polarized political climate driven by the moralization of politics (e.g. Graham et al., 2009; Haidt, 2013; Skitka & Morgan, 2014). Skitka and Morgan (2014) contrasted the ways in which previous generations viewed partisan policy differences as based on group preferences or opinions, whereas individuals in the current political moment are more likely to perceive policy issues as universal moral imperatives. Their work indicates that individuals are less likely to adopt flexible positions regarding political issues, and more likely to take extreme actions to advance their political views.

If there is indeed a widening moral divide in politics, how exactly do the two major parties fall along this divide? Are there predictable patterns for liberal and conservative morality, and if so, do Biden and Trump align with these patterns? This study explores these questions, and I turn now to a review of two prominent theories of political morality.

## Literature Review

### Morality and Politics

Cognitive linguist George Lakoff was perhaps the first to propose a model that describes the differences between liberal and conservative conceptualizations of morality. Lakoff (1996) described several observations about the ways in which liberals and conservatives approach political issues. For example, liberals and conservatives often support predictable “sets” of ideals. The liberal “set” frequently includes things like reduced military spending, support for reproductive rights (including, but not limited to, abortion), and support for environmental issues. On the other hand, conservatives often support low tax rates, increased spending on military, and strict criminal justice laws. Lakoff (1996) observed that politicians, and perhaps the public as a whole, often frame opposing viewpoints as immoral, that is, against some universally understood concept of what is right or good. As a cognitive linguist, Lakoff was interested in the differences in which liberals and conservatives talk about these moral issues, thus he developed a model of political morality, first discussed in Lakoff (1996).

Prior to developing his model of political morality, Lakoff was known for advancing Conceptual Metaphor Theory alongside Mark Johnson (Lakoff & Johnson, 1980). Broadly, this theory focuses on the ways in which people experience and describe abstract concepts in terms of concrete experiences, and how linguistic metaphors reveal these underlying concepts. For example, someone may refer to the abstract notion of *understanding* using the concrete experience of *seeing*, exemplified by metaphorical phrases such as ‘I *see* what you mean,’ or ‘My *eyes* have been *opened* to the truth.’ Following this line of reasoning, Lakoff (1996) proposed that liberals and conservatives

have distinctly different underlying conceptualizations of morals in politics. He drew on research that demonstrates a common conceptual metaphor of the United States as a *family*, with governmental entities and politicians taking *parental* roles. Lakoff (1996) proposed that the prototypical liberal understanding of this family is one led by a *nurturant parent*, while the prototypical conservative understanding is a family led by a strict parent, specifically a *strict father*. This informs how individuals understand political issues and evaluate their moral worth.

Under a pure *nurturant parent* moral concept, individuals conceptualize a family in which children are inherently good, and the role of parents is to exhibit care towards them and help them develop into mature adults. These adults then care for other people and other aspects of society, such as the environment. Nurturant parent morality follows metaphorical concepts that depict morality as empathy, nurturance, and fair distribution. In contrast, the *strict father* moral concept is based less on care and more on justice, or the idea that people should receive what they deserve based on their own initiative and action. In this family concept, children are not necessarily inherently good; they must learn good from bad and right from wrong through a just system of rewards and punishments. Strict father morality follows metaphorical concepts that depict morality as purity, health, and self reliance.

As an example of one conflict inherent between these moral concepts, Lakoff (1996) described the state of modern-day capitalism. Conservatives may point to free market capitalism to demonstrate the benefits of a just worldview. In this system, theoretically anyone can achieve wealth commensurate to the effort they put in towards this achievement. Competition in the free market begets innovation, and those who

innovate receive financial reward. Conversely, liberals may point out that this system creates a wealth gap that leaves many people behind the financial curve of elites. Those who are impoverished may not deserve their situation simply because they do not have access to the resources necessary to build wealth. Thus, the government should care for these individuals by eliminating wealth inequality in some way.

Lakoff (1996) acknowledged that his model was just that: a model, rather than a fully elaborated theory. However, he believed that these family based moral evaluations would be evident in the language used to discuss issues from either a liberal or conservative viewpoint. Researchers have since used the family model to explore the moral content of political documents, speeches, and the like, as well as to test relationships between political affiliation and family conceptualizations. Cienki (2005) analyzed debates between Republican candidate George W. Bush and Democratic candidate Al Gore during the 2000 US Presidential election. He found that each candidate did use speech that would indicate entailments of their predicted family models, but at times both candidates appeared to draw from both family models.

Deason and Gonzalez (2012) found a similar pattern when analyzing the Presidential and Vice-Presidential convention speeches from the 2008 election. The Democratic candidates Barack Obama and Joe Biden mainly drew from a nurturant parent model, but the Republican candidates John McCain and Sarah Palin drew from both a strict father and nurturant parent model. Deason and Gonzalez (2012) concluded that issue frames, rather than underlying moral concepts, appeared to be the salient factor that determined how candidates would discuss certain topics. For example, there was a significant global recession occurring at the time which warranted language from the



nurturant parent model regardless of political affiliation. It would not do to tell millions of people that they needed to pull themselves out of financial trouble, as a *strict father* moralist might; rather, candidates discussed how the government would care for citizens, in line with *nurturant parent* morality.

As a measure of individual moral leanings, Feinberg et al. (2020) described the process of developing the Moral Political Scale. This scale asks participants to rate their agreement with various statements about a family, such as “I’d rather see my children play cooperatively than competitively” or “Obedience must be instilled in children” (Feinberg et al., 2020, p. 784), with the intent of discovering a person’s underlying family concept (strict or nurturing). During development, the authors administered this scale to participants from the Amazon Mechanical Turk website alongside other scales that determined political leanings. Feinberg et al. (2020) found significant correlations between strict parent scores and indicators of political conservatism, and likewise between nurturant parent scores and indicators of liberalism. Their results supported Lakoff’s (1996) model, and, importantly, it connected this model with a newer model for morality in politics: Moral Foundations Theory (Haidt, 2013). Feinberg et al. (2020) found significant correlations between nurturant parent morality and ideals that emphasize care and fairness, as well as between strict father morality and ideals emphasizing loyalty, authority, and purity. These five categories make up the five moral foundations theorized by moral psychologist Jonathan Haidt and colleagues, a theory which I will build on for the remainder of this study.

## **Moral Foundations Theory**

Haidt (2013) and Graham et al. (2013) described the process of constructing a multidimensional theory of moral psychology by first reviewing past theories on morality that rest on only one foundation. For example, they cited Lawrence Kohlberg's promotion of a moral imperative of justice, wherein actions are deemed moral or immoral solely based on whether they promote justice, that is, whether good deeds are rewarded and bad deeds are punished. Similarly, they reviewed Carol Gilligan's work, which argued that women view morality as based on whether actions promote care towards others while preventing harm. These moral foundations are not unlike Lakoff's strict and nurturant models. Graham et al. (2013) argued that such monistic moral theories do not capture the complexity of human moral intuition because they only provide one basis on which humans judge something as moral or immoral—does an action promote justice or not? Does it promote care or not? Graham et al. (2013) described themselves as “unabashed pluralists” (p. 57), that is, their theory rests on the notion that humans have several foundations on which to draw when making moral judgments.

Haidt and Joseph (2004) began constructing a multi-foundation model of moral psychology by exploring possible evolutionary challenges of human physical and social development. This became the Moral Foundations Theory, which is comprised of five pillars of morality, each based on an evolutionary challenge of human history. Haidt and Joseph (2004) argued that humanity needed to overcome these challenges to develop within a complex social structure. The solutions to these challenges originated as social goals common to members of a social group. Common goals became group values and persist today as morals—underlying beliefs about what actions, attitudes, or behaviors are

fundamentally right or wrong (Haidt, 2013). A brief description of these foundations follows, adapted from Graham et al. (2013) and Haidt (2013):

1. The care/harm foundation, based on an evolutionary need to protect children. Current manifestations include compassion for victims of harm, both human and non-human (for example, compassion for nature or endangered species), as well as anger at the perpetrators of this harm.
2. The fairness/cheating foundation, based on an evolutionary need to form beneficial interpersonal relationships. Current manifestations include interpersonal issues, such as relational fidelity, along with non-interpersonal issues, such as equitable treatment towards all members of a society.
3. The loyalty/betrayal foundation, based on an evolutionary need to form social groups. Current manifestations include group pride ranging from school alumni and sports fans to political party and national pride.
4. The authority/subversion foundation, based on an evolutionary need to form hierarchies within social groups. Current manifestations include respect and obedience for those in authority in various contexts, such as teachers, bosses, or political leaders.
5. The sanctity/degradation foundation, based on an evolutionary need to avoid disease. Current manifestations include disgust at anything deemed a threat to personal or group sanctity, such as sexual deviancy, immigration, or physical maladies.

Graham et al. (2013) described the process of developing Moral Foundations Theory as the co-development of theory and methodology. During the early stages of

development, the theory informed researchers on methods to test the theory; these methods then informed new directions and revisions for the theory, spurring new methods and so forth. Moral Foundations Theory has primarily been tested in four categories of methodology: word count analyses, self-report scales, implicit measurements, and physiological measurements. I describe some of the studies here with the exception of the word count analyses, to which I devote more attention in upcoming sections.

### *Self-report scales*

Graham and Haidt (2012) described the development and administration of a Moral Foundations Sacredness Scale. This scale asked participants how much they would need to be paid to commit a moral violation, such as burning the national flag to measure loyalty or “kicking a dog in the head” (Graham & Haidt, 2012, p. 22) to measure care. Participants could respond on a scale ranging from “I’d do it for free” to “not for any amount of money” with a range of monetary values between these extremes from \$10 to one million dollars. Results showed that conservatives were less likely than liberals to accept a monetary tradeoff for violations of the loyalty, authority, and sanctity foundations. However, both liberals and conservatives were equally likely to refuse a tradeoff for the care and fairness foundations.

### *Implicit Measures*

In a “foundation tradeoff” task, Graham (2010) addressed the concern that moral foundation research focuses on isolated moral content, whereas real world moral judgments are often enmeshed in contexts that involve decisions between different circumstances. After completing a questionnaire to determine political leanings,

participants were asked a series of tradeoff questions to choose between the “morally worse” of two different potential moral violations, for example “running a red light vs. frowning at a stranger” (Graham, 2010, p. 23). Participants were assigned to either a condition that encouraged quick gut-level reactions or a condition that encouraged slow deliberation before making a choice. Across both conditions, those participants who identified as liberal favored the care and fairness foundations; that is, when faced with a tradeoff of moral violations, liberals consistently rated those which involved harm or cheating as the worse violation. Conservatives rated violations of the loyalty, authority, and sanctity foundations as worse when paired against care or fairness violations.

These results of the foundation tradeoff test are consistent with self-report measures of moral foundation endorsement, although Graham (2010) acknowledged that this test only begins to explore the possibility of implicit versus explicit judgment. Graham (2010) expanded on this by performing affective priming tests, wherein participants judged neutral words as positive or negative based on a particular priming word. In one test, participants at a computer screen saw a “vice” word, that is, a word that violates one of the five moral foundations, such as “kill.” This word flashed for 150 milliseconds on a computer screen, acting as a prime. Participants then rated a neutral word as either positive or negative, with the assumption that a negative rating would indicate that the participant reacted with stronger emotional valence after seeing the prime word. Participants who identified as politically liberal rated the neutral words paired with a ‘fairness’ vice as negative more often than conservative participants, suggesting that liberals place greater emphasis on this foundation when making intuitive moral judgments. Results across other foundations were similar to the foundation tradeoff

task, with conservative participants placing greater emphasis on the loyalty, authority, and sanctity foundations.

### ***Physiological Measures***

Cannon et al. (2011) measured facial micro expressions to explore the link between emotions and morality. Participants were presented with various scenarios representing violations of the five moral foundations, such as “someone cheated on a game of cards” violating the fairness foundation or “someone gossiped about a friend” (Cannon et al., 2011, p. 327) violating the loyalty foundation. Participants were asked to give an appraisal of the scenario on a seven-point scale ranging from very negative to very positive. Unbeknownst to the participants, researchers also recorded their facial expressions during the study, with these facial expressions indicating affective responses of disgust, anger, or unspecified positive affect. The researchers found a significant correlation between affective response and moral appraisal, such that the strongest affective responses were associated with the lowest appraisal ratings (i.e., those rated as most negative). Cannon et al. (2011) did not include political affiliation as part of their analysis, but their results do support the assertion of a multi-foundation model of morality.

### ***Moral Foundations and Politics***

Graham et al. (2013) pointed out that Moral Foundations Theory was not initially developed as a theory of political morality, but throughout development researchers consistently noted that a partisan divide appeared to exist within the five foundations. For example, Graham et al. (2009) found that study participants who completed the moral foundations questionnaire and identified as politically liberal more often endorsed moral

issues based in the care and fairness foundation, while conservative participants endorsed moral issues based in loyalty, authority, and sanctity. Haidt (2013) suggested that this divide is due, at least in part, to the individuating aspects of the first two foundations compared to the conforming aspects of the remaining three. The care and fairness foundations focus on individuals as moral agents, which may correspond to progressive political philosophies that place emphasis on individual liberties. This is opposed to the loyalty, authority, and sanctity foundations, which focus on groups as the moral agents and may correspond to conservative philosophies that emphasize group conformity.

To explore the political divide of Moral Foundations Theory, Feinberg and Willer (2015) measured participant reactions to various political arguments framed by language that would align the argument with either typically liberal moral foundations or typically conservative ones. In one example, they framed an argument for universal health care in terms of basic human rights (appealing to the care and fairness foundations) or in terms of physical health and purity (appealing to the sanctity foundation). Similar questions were used based on English as a national language or the legality of same-sex marriage. Across their tests, the authors found that participants were more likely to support an issue based on the moral framing of that issue, rather than the issue itself. If the issues were framed in a way that appealed to typically conservative moral foundations, conservative participants were more likely to claim that they supported the issue, and likewise for liberal participants.

Participants were then prompted to write arguments for policy issues with the explicit goal of persuading someone of opposing political affiliations (Feinberg & Willer, 2015). Participants consistently wrote these arguments in a way that appealed to their

own moral foundations, rather than the anticipated foundations of their audience. Taken alongside the authors' initial findings that the moral framing of a message impacts acceptance more than the message itself, the participants' written arguments would likely be ineffective in persuading their intended audience. In addition, this suggests that people are generally unaware of the moral divide between parties and how their own moral reasoning may bias attempts at communicating across this divide.

It appears then, that American politics is couched in moral reasoning. If a politician attempts to persuade his or her audience with morally charged language, it is possible that they will only succeed in persuading those who already align with the moral convictions referenced, as suggested by Feinberg and Willer (2015). This could be one factor as to why the 2020 presidential election was so evenly divided, perhaps because the two primary candidates failed to reframe their messages for those with differing moral foundations. Moral Foundations Theory provides one lens through which to examine moral content of political messages. The theory suggests that the candidates' language should reference their moral positions in predictable ways according to political party. Biden's language should reference the individuating moral foundations of care and fairness more often than Trump's language. Similarly, Trump's language should reference the conforming foundations of loyalty, authority, and sanctity more often than Biden's language. Thus, I propose five hypotheses to test while analyzing the content of each candidate's campaign speeches:

H1: Biden's speeches will contain greater moral content from the care foundation than Trump's speeches.



H2: Biden's speeches will contain greater moral content from the fairness foundation than Trump's speeches.

H3: Trump's speeches will contain greater moral content from the loyalty foundation than Biden's speeches.

H4: Trump's speeches will contain greater moral content from the authority foundation than Biden's speeches.

H5: Trump's speeches will contain greater moral content from the sanctity foundation than Biden's speeches.

One way to test these hypotheses is to use a keyword analysis based on a moral foundations reference word list. This type of analysis yields scores for each text analyzed indicating the moral content found under each of the five moral foundations. Further statistical analysis can identify if the mean scores of each category differ between candidates at a statistically significant level. To date, three Moral Foundations Dictionaries have been developed. The first two came out of research by some of the originators of Moral Foundations Theory, while a third was developed independently. I describe the process for each below.

## **Defining and Measuring Moral Content**

### ***The Moral Foundations Dictionary***

Graham et al. (2009) outlined the first attempts to use word count analysis to identify the moral content of a text. The authors began by developing a reference list of words for each moral foundation under the assumption that a speaker's use of certain words would reveal his or her underlying moral conceptualizations. The resulting word list was dubbed the Moral Foundations Dictionary (MFD). Initial development of this

dictionary involved a group that included the authors and five research assistants. This group began with the base words for each foundation (i.e., care, harm, fairness, cheating, etc) and then used dictionaries and thesauruses to find synonyms and other words related to these bases. This led to a large list of possible keywords for each moral foundation. The authors then reduced the list of words to eliminate those that did not seem directly relevant to the corresponding foundation as well as words that carried multiple meanings not related to morality (e.g. ‘just’ means ‘fair’ but also ‘only’). The resulting Moral Foundations Dictionary used by Graham et al. (2009) contains 295 total words and word-stems (e.g. “kill\*” as a root for ‘killer,’ ‘killing,’ ‘killed,’ etc.) across the five moral foundations.

To test the validity of the moral foundations dictionary, Graham et al. (2009) used a corpus of religious sermons based on the assumption that these would naturally contain a high level of moral content and that this content would fall along predictable lines of liberal versus conservative content based on the religious source. Graham et al. (2009) reviewed research that correlated political leanings and political involvement with various Christian denominations. This led them to conclude that Southern Baptist sermons and Unitarian Universalist sermons should give clear examples of conservative and liberal moral content, respectively. In line with the broader Moral Foundations Theory, Graham et al. (2009) hypothesized that the Southern Baptist sermons would contain more words from the loyalty, authority, and sanctity foundations, while the Unitarian Universalist sermons would contain more words from the care and fairness foundations.

To test their hypotheses, Graham et al. (2009) used the Linguistic Inquiry and Word Count (LIWC) computer program. This program uses a reference dictionary (in this case, the moral foundations dictionary) to analyze a set of text files. It returns scores for each text file from each category defined in the reference dictionary. This score indicates moral keyword density, measured as the percentage of total words in the analyzed text file that are present in each moral foundation. For example, a keyword density of .11 in the “care” category of a 1,000-word file would indicate that .11% of those 1,000 words (or eleven words total) matched an entry for “care” in the reference dictionary.

The basic assumption of this form of analysis is that words that appear in one moral foundation with greater frequency than others may indicate that the author or speaker of the analyzed text is endorsing that foundation over the others. Graham et al.’s (2009) initial results mostly supported their hypotheses, with the liberal sermons containing greater moral content from the care and fairness foundations when compared to the conservative sermons, while the conservative sermons contained greater moral content from the authority and sanctity foundations compared to the liberal. However, the comparison for the loyalty foundation was not in the direction predicted; words appeared here more frequently in the liberal sermons.

Graham et al. (2009) considered the likelihood that context may have played an important role in this difference. While a computer can tally words based on a list of reference moral foundational words, it cannot determine how the word is being used in context. For example, the word may be used to either support or deny the category, it may be used ironically or as part of a quote. Instances such as these could negate the

validity of the analysis. To correct for this possibility, Graham et al. (2009) included a post-hoc contextual analysis of the sermons, wherein human coders used the output of the Linguistic Inquiry and Word Count analysis to validate or adjust the initial results. The authors took a sample of the words highlighted along with two or three sentences of surrounding context to create clusters of keywords plus context. After scrambling these clusters and removing information that would indicate their source texts, four respondents scored each cluster on whether the moral keyword appeared to support or deny the assigned moral category, or indeed whether it appeared to be neutral overall. After this adjustment took place, the results for the loyalty foundation reversed. The coders found that the liberal sermons most often used words from this foundation to criticize, rather than support, loyalty-based morality. The contextual analysis did not reverse results for any other moral foundations, thus lending overall support to the validity of the word count analysis with the caveat that further analysis may be beneficial in interpreting the results.

The reliance on human-based contextual analysis to correct automated data represents a potential limitation of automated word-count analysis, that this automation may not be able to reveal the whole story of moral content in an analyzed corpus. This is particularly problematic for Moral Foundations Theory because it describes morality as an inherently intuitive process. The moral foundations dictionary was developed by a small group of researchers who deliberated over which words should be included in the dictionary. Such deliberation may be at odds with the underlying concept of morality being an intuitive, rather than deliberate, process. Since Graham et al.'s (2009) initial work, researchers have sought to address this limitation with two variants of the moral

foundations dictionary: the Moral Foundations Dictionary 2.0 (Frimer et al., 2017) and the extended Moral Foundations Dictionary (Hopp et al., 2020). Both were developed with the goal of moving beyond a small list of deliberately chosen words, though each took a different path to this end. I describe the two processes below, beginning with the Moral Foundations Dictionary 2.0.

### ***The Moral Foundations Dictionary 2.0***

Frimer et al. (2017) reviewed Graham et al. (2009) and commented on two primary limitations of the Moral Foundations Dictionary, the relatively short list of words in the dictionary and the lack of certain words that appear to be prototypical of each domain (such as ‘murder’ or ‘torture’ in the care/harm foundation). To address this, Frimer et al. (2017) first generated a large list of words for each moral foundation in a manner not unlike the initial “deliberation by experts” described by Graham et al. (2009). However, to move beyond this deliberative process, Frimer et al. (2017) analyzed this new list of words using the Word2Vec computer program to validate which words were highly prototypical of each moral foundation<sup>1</sup>. Word2Vec returned a vector score for each word in the new word list that indicated how likely each word would be found alongside other foundational words in natural language. From these scores, Frimer et al. (2017) selected those above a statistically significant threshold, resulting in a total word list of 2,103 words, nearly ten times as many as the initial Moral Foundations Dictionary.

---

<sup>1</sup> Word2Vec, developed by Google, uses artificial intelligence to analyze an input text and compare it to a large database of reference corpora that is meant to capture natural language use. Word2Vec assigns numerical scores to words in the reference corpus based on how often they appear in context with words in the input text. Words with high scores indicate a higher likelihood that they are used alongside a reference word. Thus, each input word receives a “vector” score of related words.

The second step in creating the Moral Foundations Dictionary 2.0 was to test the validity of the new dictionary alongside the initial dictionary developed by Graham et al. (2009). Frimer et al. (2017) recruited participants from a crowdsourcing website, 656 of which completed an essay writing task. After reading information about a randomly chosen moral foundation, each participant was instructed to write about a time in which someone acted either in accordance with or against that foundation. This created a corpus of natural language texts which Frimer et al. (2017) analyzed in the Linguistic Inquiry and Word Count program with their new moral foundations dictionary. By analyzing a text that was specifically written about someone exhibiting care, for example, the authors explored whether their new moral dictionary would detect moral content in that category at a higher rate than the previous dictionary. This proved to be the case, with higher keyword density scores and higher effect sizes across all categories when compared to the Moral Foundations Dictionary.

Both the Moral Foundations Dictionary and the Moral Foundations Dictionary 2.0 used word lists that reflect an implicit operationalization of the moral content of a textual corpus. Specifically, morality in a text is based on language as it appears to be used, built from lists based on the assumptions of word usage to reveal underlying moral thought. The measurement of this language in terms of keyword density is an attempt to describe the moral content of a text in terms of the author's use of certain words and, by possible extension, the intent of the author to convey a particular moral message. Thus, the focus appears to be on the author of a text, despite the fact that an author's intent to communicate a particular moral message may not be the same message that an audience interprets. Researchers in the Media Neuroscience Lab at the University of California,

Santa Barbara developed a new method for measuring moral content that addresses this angle, as well as other potential shortcomings of Frimer et al.'s (2017) method.

### ***The extended Moral Foundations Dictionary***

Whereas the first Moral Foundations Dictionary (Graham et al., 2009) relied on a small group of people with subjective interpretations of which words would likely fall into each moral category, the updated version 2.0 is built on mathematical models of natural language use to increase construct validity, followed by an analysis of participant written texts to test this validity. However, Hopp et al., (2020) noted that the Moral Foundations Dictionary 2.0 is still primarily based on expert deliberation. The mathematical modelling performed by the Word2Vec program used a deliberative set of seed words to begin with, and so the resulting dictionary may still be constrained in its ability to reflect the intuitive process of moral interpretation. Hopp et al. (2020) addressed this by creating a new moral dictionary, the extended Moral Foundations Dictionary. It is still based on Moral Foundations Theory but developed from the ground up in a process different from the first two iterations.

The extended Moral Foundations Dictionary (Hopp et al., 2020) takes a crowdsourced approach to dictionary construction. Rather than assuming a list of words and testing whether they fit into a specific category, the extended Moral Foundations Dictionary was built with a greater dependence on a priori human input, essentially front-loading the dictionary with participant-rated interpretations of moral keywords. This was done via a crowd-sourced annotation task, wherein 557 participants from across the United States read 20 articles each from a sample of 2,995 newspaper articles. These participants produced 63,958 annotations, indicating words and phrases that they felt

represented an appeal to any or all of the five foundations described by Moral Foundations Theory. From these annotations, Hopp et al. (2020) applied various filtering and data cleaning tasks to extract a final quantity of 3,270 words from these annotations, each of which was assigned a score from 0 to 1 for each of the five moral foundations. These scores represent the number of times a participant annotated that word for a moral foundation divided by the total number of times the word was seen by any participant.

In addition to developing and implementing the methods to build a new moral foundations dictionary, Hopp et al. (2020) developed a companion computer program to perform analyses with their dictionary. The program, eMFDScore, is available as open source software from the lead author's GitHub repository (Hopp et al., 2019/2021).

eMFDScore is similar to LIWC2015 in that it can analyze a corpus based on a given reference list, although a researcher is limited to using either the first two Moral Foundations Dictionaries or the new extended Moral Foundations Dictionary.

Theoretically then, it is possible to compare analyses using different dictionaries within one program, and Hopp et al. (2020) did just that. There do seem to be some limitations in these comparisons, but before addressing these it is important to consider what is being measured when using the extended Moral Foundations Dictionary and how the results are to be interpreted.

The foundation scores returned by eMFDScore when using the extended Moral Foundations Dictionary indicate “the probability that a particular word was annotated with a particular moral foundation” (Hopp et al., 2020, p. 237). The authors describe the computation of these scores as the number of times a participant associated a word with a particular foundation divided by the number of total participants that saw that particular



instance of the word. The resulting statistic indicates the proportion of annotations that associated a word with each of the foundations. However, there does not appear to be anything probabilistic about this. Scores based on the extended Moral Foundations Dictionary are based on a past event that can be measured with certainty, that is, the annotation procedure that had definite, measurable outcomes.

The discussion of “probability” versus “proportion” may just be an issue of semantics, and it is not my intent to argue against the exact verbiage used. However, this verbiage reflects a potentially problematic conceptual basis for the measure. If the statistics returned by eMFDScore are indeed probabilities, this seems to indicate that they may be used as inferential statistics, rather than the strictly descriptive style of a LIWC analysis, for example. Yet this raises the question of what exactly a researcher can infer with these probabilities? If a document receives a “mean probability” score of .15 for the care foundation, does this mean that there is a 15% chance that any given member of the public will interpret a message related to care? Is it simply a 15% chance that another round of the same annotation procedure would interpret a message related to care?

The best that Hopp et al. (2020) have to offer is an explanation that their scoring system reflects a proportion of interpretation from their initial participant pool, yet they do not specify precisely what this statistic might mean when applied to the analysis of any given set of texts. If their participant pool is meant to reflect the interpretation of the general public, the authors do not justify what makes their pool representative of a wider population. Haidt (2013) described in depth how the moral foundations appear to be consistent across cultural bounds. It is thus difficult to accept that a relatively small

sample of participants based in the United States would form an adequately representative pool that can reliably predict the moral intuitions of humanity at large.

### ***Comparing the Methods***

Hopp et al. (2020) discussed the superiority of the extended Moral Foundations Dictionary compared to the Moral Foundations Dictionary and Moral Foundations Dictionary 2.0. The authors analyzed the same corpus with all three dictionaries to demonstrate that the extended Moral Foundations Dictionary returned higher scores overall for each moral foundation and that these scores were more normally distributed than the previous two dictionaries. Further, they showed that significant positive correlations existed between scores for the same moral foundations across dictionary types. Thus, they concluded that the extended Moral Foundations Dictionary presents a better understanding of the moral content of a text, and that the correlations between dictionaries show that the methods were detecting the same “moral signal” (Hopp et al., 2020, p. 240).

However, there appear to be some major limitations to this comparison. While Hopp et al. (2020) used the Moral Foundations Dictionary and Moral Foundations Dictionary 2.0 in their original forms, they did not score their target documents in the same way that previous research on these dictionaries has done (e.g. Graham et al. 2009, Frimer et al. 2017, Frimer 2019). The eMFDScore program can use the Moral Foundations Dictionary and Moral Foundations Dictionary 2.0 dictionaries, but it scores items by computing the proportion of total moral keywords identified that exist within each foundation. For example, if an item contains 100 total words found anywhere in the Moral Foundations Dictionary with 24 in the care foundation specifically, eMFDScore

returns a score of .24 for the care foundation. This is regardless of the overall wordcount of the item, which previous methods used to calculate keyword density of a text. Further, when using eMFDScore to analyze a corpus with the extended Moral Foundations Dictionary itself, the scoring again changes. Rather than calculating each moral foundation as a proportion of total moral content, eMFDScore returns an average of the scores assigned to words in the extended Moral Foundations Dictionary. It is unclear then how these scores can be directly compared or why, for instance, “higher scores” using the extended Moral Foundations Dictionary indicate anything substantial about its superiority over other dictionaries.

### ***Summary***

The first two iterations of the Moral Foundations Dictionary were based on deliberation about what words would likely indicate an appeal to a given moral foundation. This inherently creates dictionaries that describe morality based on words as they are *used*, or as they are presumed to be used. The validation test by Frimer et al. (2017) took this one step further by analyzing documents to determine how a participant pool used words in relation to morality. However, the extended Moral Foundations Dictionary was created based on how a participant pool *interpreted* words in relation to morality. Thus, the first methods define moral content based on the presumed intended use of language, while the more recent method defines it based on the apparent interpretation of language.

In addition to inherent differences surrounding the definition of moral content, there are distinct differences in the quantification of moral content depending on the analysis program used. Assuming all methods are adequately reliable and valid, it may be

possible for a researcher to use more than one method to triangulate his or her understanding of the morality of a text. Aside from this, a direct one-to-one comparison of methods to determine the superiority of one over all does not seem warranted; the methods simply do not measure the same concept in the same way.

It is not my intent to dismiss the work of Hopp et al. (2020). Indeed, they gave a compelling rationale for creating a crowd-sourced dictionary with a continuous scoring scheme across moral foundations. Human morality is nuanced and context specific, and their method may reflect this more accurately than the discrete, binary counting method provided by a LIWC analysis. Therefore, it seems natural to conclude that the extended Moral Foundations Dictionary offers greater inferential power than previous method. However, without additional tests to solidify the validity and reliability of the method in different contexts, any insight derived from an eMFDScore analysis should not be accepted as generalizable to a broad audience. Because Hopp et al. (2020) did not make a clear argument as to what their analysis results truly indicate, I tested my hypotheses with a LIWC2015 analysis using the Moral Foundations Dictionary 2.0. That said, I was still interested in how an eMFDScore analysis would compare, and so I performed this analysis as well.

## Experiment 1

### Method

#### *Sample*

I used a purposive sampling strategy to include all instances of public speaking, such as speeches, debates, and town halls, given by the Presidential candidates between August 17, 2020 (the first day of the Democratic National Convention) and November 4, 2020 (Election Day). This sample contains 131 events in total, of which 56 involve Joe Biden and 75 involve Donald Trump. I have chosen this specific period of the election season because it involves direct campaigns between two candidates of opposing political parties. Events earlier in the season may have included campaigning against others within the same party, for example during the primary elections. While it would be interesting to explore differences in moral content within the same party, that is beyond the scope of the proposed research.

I obtained full transcripts of the selected events from the website Rev.com, which uses software to create automated transcriptions of a wide variety of sources. The website offers free transcripts of political speeches as a public service. Users can watch a speech on Rev.com while following along with the transcription, thus allowing a quality check of passages that appear unclear or incorrectly transcribed. I used this function along with Microsoft Word's "Find and Replace" feature to edit out unclear or extraneous information, such as words from other speakers, speaker identification markers, timestamps, and so on. Thus, I ended with 131 text files containing nothing more than each speakers' words.

Biden's sample (N=56) ranged in word count from 408 words to 10,587 words (M=3,457.45, SD=1,725.36), while Trump's sample (N=75) ranged in word count from 1,138 words to 18,303 words (M=9,292.25, SD=3,930.14). This reveals a difference in mean word count of nearly 6,000 words, and a *t*-test shows that this difference is statistically significant, with  $t(129)=10.380$ ,  $p<.001$ . This difference makes some sense when comparing the contexts of each speakers' events. Most of Trump's speeches were delivered at large rallies, during which Trump was known to speak extemporaneously about various topics. In contrast, Biden's events were smaller in scale, and he may have been more likely to follow a close script.

These discrepancies cannot be ignored when running a word count analysis. However, using overall word proportions rather than raw word counts allows for an easier comparison between situations with drastically different word counts. If a speaker consistently draws from a particular moral foundation, then there would presumably be the same proportion of these moral keywords despite the speaker's overall word count. This assumption underlies many of the studies that led to the development of Linguistic Inquiry and Word Count itself (Tausczik & Pennebaker, 2010) as well as the first two iterations of the Moral Foundations Dictionary (Graham et al., 2009; Frimer et al., 2017; Frimer, 2019).

### ***Design***

This first analysis tested the moral content of my sample to detect differences predicted by Moral Foundations Theory. The independent variable was the speaker, either Biden or Trump. The unit of analysis was an individual speech, analyzed as a single text file per speech. The dependent variables were the measure of moral content

for each moral foundation, operationalized here as the density of moral keywords present in each speech, that is, the proportion of individual speech word count that falls under each moral foundation.

### ***Procedure***

I used the Linguistic Inquiry and Word Count computer program (version LIWC2015) alongside the Moral Foundations Dictionary 2.0 (Frimer et al., 2017). I separately analyzed the sample of Biden's speeches and the sample of Trump's speeches to obtain scores for the five moral foundations. These scores reflect the proportion of overall words in the analyzed speech that fall under each category. For example, a speech of 100 words that contains five words in the care foundation would receive a score of 5.00 for that foundation. While the Moral Foundations Dictionary 2.0 allows separate analysis for "virtue" and "vice" words for each category, I combined these for overall category analysis. My hypotheses do not directly address instances of "virtue" versus "vice" words, and the combination of these counts to reflect a total foundation score is in line with previous studies using the moral dictionaries (e.g., Frimer, 2019; Graham et al., 2009; Hopp et al., 2020). Thus, the LIWC2015 program returned overall word count for each item, total keyword density across all moral foundations, and total keyword density for each foundation individually. I loaded this data into IBM SPSS version 28 for an ANOVA of the five categories to test my hypotheses at the significance level of  $p < .05$ .

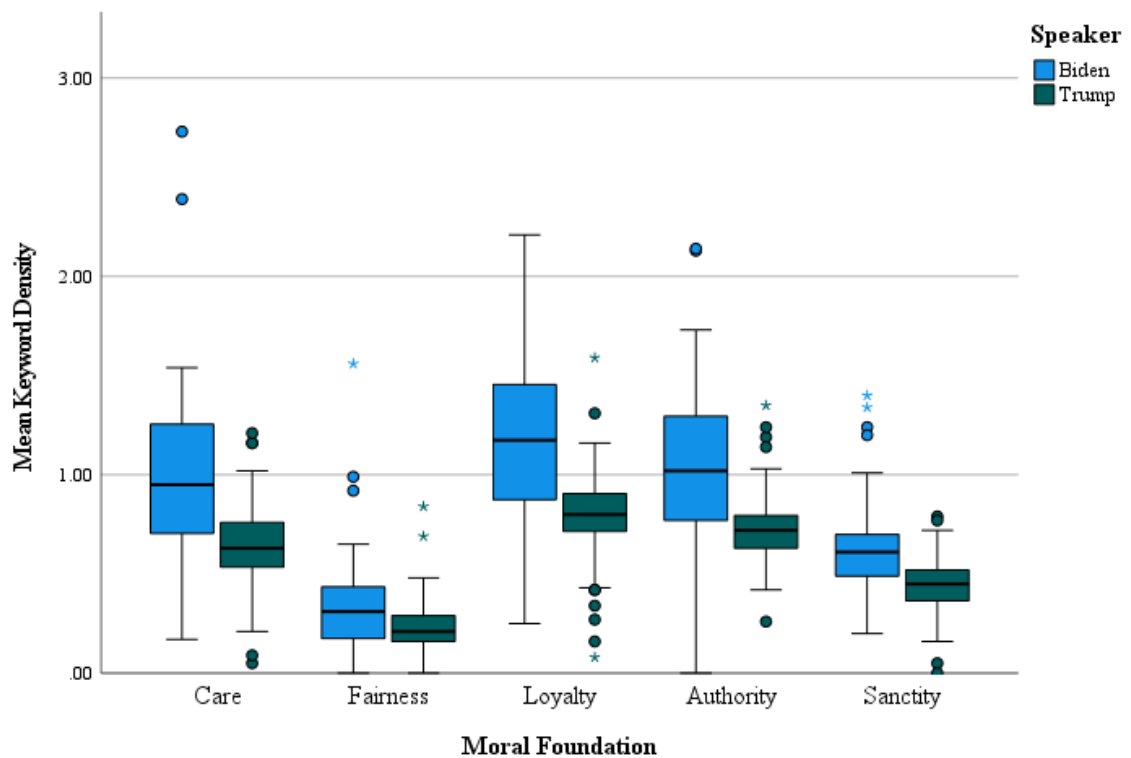
### **Results**

Figure 1 shows a boxplot of each candidate's moral content based on mean moral keyword density per speech. Based on this boxplot and initial descriptive statistics (see table 1), it appeared that Biden's mean scores are higher across all five moral foundations

than Trump's scores. Before proceeding with the ANOVA to test these differences, I first checked the skewness of each category. A common critique of word count analyses is that the results are often highly skewed (e.g., Frimer, 2019; Hopp et al., 2020). With the slight exception of Biden's "fairness" foundation that showed a skewness statistic of 2.28 (see table 1), my analysis appeared to be fairly regular overall. Frimer (2019) applied a logarithmic transformation to his MFD analyses to correct for skewness, but this did not appear to be advantageous in my case.

**Figure 1**

*Boxplot Depicting Moral Keyword Analysis Using LIWC2015 and the Moral Foundations Dictionary 2.0.*





**Table 1**

*Descriptive Statistics of LIWC2015 Analysis Using the Moral Foundations Dictionary*

2.0.

	Care	Fairness	Loyalty	Authority	Sanctity
Biden <sup>a</sup>					
Min	.17	.00	.25	.00	.20
Max	2.73	1.56	2.21	2.14	1.40
M	.99	.35	1.19	1.04	.62
SD	.47	.26	.43	.41	.25
Skew	1.11	2.28	.19	.32	1.16
Trump <sup>b</sup>					
Min	.05	.00	.08	.26	.00
Max	1.21	.84	1.59	1.35	.79
M	.64	.24	.79	.72	.44
SD	.22	.13	.23	.19	.15
Skew	-.13	1.84	-.24	.79	-.39

<sup>a</sup> $n = 56$ . <sup>b</sup> $n = 75$ .

**Table 2**

*Means, Standard Deviations, and One-Way Analysis of Variance of Moral Content Using LIWC2015 and the Moral Foundations Dictionary 2.0.*

Moral Foundation	Biden		Trump		$F(1, 129)$
	$M$	$SD$	$M$	$SD$	
Care	.99	.47	.64	.22	32.15***
Fairness	.35	.26	.24	.13	11.36***
Loyalty	1.19	.43	.79	.23	47.20***
Authority	1.04	.41	.72	.19	35.73***
Sanctity	.62	.25	.44	.15	27.50***

\*\*\*  $p < .001$ .

Results of the ANOVA are shown in table 2. Across all five moral foundations, there is a statistically significant difference between Biden's and Trump's moral content as measured by moral keyword density, or the proportion of overall word count found in each moral foundation. Regarding the care foundation, Biden's moral content ( $M=.99$ ,  $SD=.47$ ) was significantly higher than Trump's moral content ( $M=.64$ ,  $SD=.22$ ),  $F(1, 129)=32.15$ ,  $p<.001$ ; this supports H1.

Regarding the fairness foundation, Biden's moral content ( $M=.35$ ,  $SD=.26$ ) was significantly higher than Trump's moral content ( $M=.24$ ,  $SD=.13$ ),  $F(1, 129)=11.36$ ,  $p<.001$ ; this supports H2.

Regarding the loyalty foundation, Biden's moral content ( $M=1.19$ ,  $SD=.43$ ) was significantly higher than Trump's moral content ( $M=.79$ ,  $SD=.23$ ),  $F(1, 129)=47.20$ ,  $p<.001$ . While this result is significant, the direction of difference between candidates is counter to my prediction in H3.

Regarding the authority foundation, Biden's moral content ( $M=1.04$ ,  $SD=.41$ ) was significantly higher than Trump's moral content ( $M=.72$ ,  $SD=.19$ ),  $F(1, 129)=35.73$ ,  $p<.001$ . While this result is significant, the direction of difference between candidates is counter to my prediction in H4.

Regarding the sanctity foundation, Biden's moral content ( $M=.62$ ,  $SD=.25$ ) was significantly higher than Trump's moral content ( $M=.44$ ,  $SD=.15$ ),  $F(1, 129)=27.50$ ,  $p<.001$ . While this result is significant, the direction of difference between candidates is counter to my prediction in H5.

### ***Within-Subject Comparisons***

While performing the above analysis, I observed that the overall pattern of moral content per foundation appeared to be similar for each candidate. That is, each candidate appeared to emphasize words from the loyalty foundation, followed by the authority foundation, the care foundation, and the sanctity foundation, with the least amount of emphasis on the fairness foundation. To test whether the differences between categories are significant, I ran pairwise t-tests within each candidate's corpus. I began with the loyalty foundation as the category with the highest mean scores for both candidates,

compared with the authority foundation with the second highest mean scores. Subsequent pairs followed this ranking pattern. Results are displayed in table 3.

**Table 3**

*Within-Subject Comparisons Using LIWC2015 and the Moral Foundations Dictionary*

2.0.

	Biden				Trump			
	M	SD	<i>t</i> (55)	<i>p</i>	M	SD	<i>t</i> (74)	<i>p</i>
Pair 1								
Loyalty	1.19	.43	2.12	.038	.79	.23	2.54	.013
Authority	1.04	.41			.72	.19		
Pair 2								
Authority	1.04	.41	.75	.458	.72	.19	3.19	.002
Care	.99	.48			.64	.22		
Pair 3								
Care	.99	.48	6.57	<.001	.64	.22	9.57	<.001
Sanctity	.63	.25			.44	.15		
Pair 4								
Sanctity	.63	.25	6.25	<.001	.44	.15	7.90	<.001
Fairness	.35	.26			.24	.13		

*Note.* All *p* values reflect two-tailed *t*-tests.

## Discussion

Initial analysis of the moral content of my sample using LIWC2015 and the Moral Foundations Dictionary 2.0 provided mixed support for my hypotheses. Biden's moral content as measured by keyword density surpassed Trump's content for all moral foundations, not just those predicted by Moral Foundations Theory. Using this measure alone, one could find support for the observation that Biden's goal was to be a 'values candidate' who emphasized a return to morality in contrast to the previous administration. Such a candidate may very well use a higher instance of morally salient language overall. However, results also revealed similar patterns of moral keyword use between the candidates. Except for a statistically non-significant difference between

Biden's mean keyword usage for 'authority' and 'care,' both candidates essentially emphasized the same moral foundations within their speeches.

Analysis reveals statistically significant differences when comparing means between candidates. However, given their apparent similarities in moral orientation, or the pattern of moral foundation emphasis, there may be no real substantive differences between them. It may simply be the case that Trump gave longer speeches with more neutral content, but when it comes down to moral language, perhaps both candidates were more similar than they were different. This reveals a potential stumbling block when using a keyword counting analysis such as the Linguistic Inquiry and Word Count program. A researcher could draw very different conclusions depending on both how a study operationalizes the underlying concept of interest and whether tests are carried out between or within subjects.

It is worth considering the overall word count per foundation of the Moral Foundations Dictionary 2.0 when evaluating these results. Perhaps the scores for 'fairness' and 'sanctity' were low because the reference list was simply smaller than the other foundations? This does not appear to be the case. The Moral Foundations Dictionary 2.0 contains an average of 421 combined virtue and vice words for each moral foundation. The foundation with the smallest number of words is loyalty with only 192 words, yet both candidate's highest mean scores were in this category. The fairness foundation contains 351; still below average, but seemingly not enough to warrant scores significantly lower than the remaining foundations. Interestingly, the sanctity foundation contains the highest word count at 660, yet this foundation ranked second to last in the analysis.

## Experiment 2

### Method

I wanted to know how these results compared to an analysis using the extended Moral Foundations Dictionary while avoiding Hopp et al.'s (2020) apparent gap in comparing these methods. If I were to run an analysis using the eMFDScore program, it would not provide a direct comparison to my initial results given the inherent differences in score calculation and overall operationalization. I needed to take a few preliminary steps before performing the eMFDScore analysis: reformatting the corpus for use with eMFDScore and performing an eMFDScore analysis using the Moral Foundations Dictionary 2.0. I describe these processes below.

### *Testing a Segmented Corpus*

First, I adjusted the corpus to reflect the requirements for eMFDScore, which vary from LIWC2015. EMFDScore requires a corpus to be compiled into a single spreadsheet document with every item entered into a single cell. I created one spreadsheet for each candidate in Microsoft Excel, but I encountered a software limitation in the process. A single cell in Excel is limited to 32 kilobytes of information, or roughly 6,000 words. This was not a problem for most of Biden's speeches, in which only two items exceeded this limit and required me to split the item over two cells. However, I was required to split many of Trump's speeches into two or three cells. This meant that, rather than analyzing 131 total items, I would now be analyzing 211 total items. See table 4 for descriptive statistics of the reformatted corpus.

**Table 4**

*Descriptive Statistics of LIWC2015 Analysis Using the Moral Foundations Dictionary 2.0 with Segmented Corpus.*

	Care	Fairness	Loyalty	Authority	Sanctity
Biden <sup>a</sup>					
Min	.17	.00	.25	.00	.00
Max	2.73	1.56	2.21	2.14	1.40
M	.97	.36	1.16	1.02	.60
SD	.48	.27	.45	.41	.27
Skew	1.04	2.06	.12	.31	.87
Trump <sup>b</sup>					
Min	.05	.00	.08	.00	.00
Max	1.83	.84	4.18	1.47	1.82
M	.72	.23	.93	.73	.49
SD	.33	.14	.52	.23	.22
Skew	.92	1.72	3.38	.29	1.89

<sup>a</sup> $n = 59$ . <sup>b</sup> $n = 152$ .

Biden's sample ( $N=59$ ) ranged in word count from 408 words to 6,118 words ( $M=3,281.64$ ,  $SD=1,262.03$ ), while Trump's sample ( $N=152$ ) ranged in word count from 223 words to 6,258 words ( $M=4,639.83$ ,  $SD=1,848.22$ ). This revealed a much smaller difference in mean word count than the initial analysis, though a  $t$ -test showed that this difference is still statistically significant, with  $t(209)=5.19$ ,  $p<.001$ . The new corpus not only has a different sample size, but essentially a different unit of analysis, moving from 'entire speech' to simply a speech segment in the case of larger speeches. It is possible that this change alone could lead to differences in analysis. If so, then I would have less confidence in comparing the analyses between LIWC2015 and eMFDScore, since there would be fundamental differences in the corpus.

Given that only two of Biden's speeches needed to be split for reformatting, his moral foundation statistics did not change much. However, Trump's statistics are more skewed, particularly within the loyalty foundation. This is perhaps due to certain

segments of Trump’s speeches containing a higher density of loyalty content that essentially became “diluted” when examining the speeches as a whole. Though given that Trump’s overall moral content is not significantly skewed, I did not apply any adjustments before running the ANOVA.

I analyzed this new corpus with LIWC2015 as in Experiment 1 under the assumption that similar results would lend confidence in a comparison between the two methods. Results of this ANOVA were in line with my initial analysis of the unmodified corpus of speeches; see table 5 for full results. Biden’s speeches (and speech segments) still contained a higher mean moral keyword density than Trump’s, though results were slightly less statistically significant for the loyalty foundation (changed from  $p<.001$  to  $p=.003$ ) and the sanctity foundation (changed from  $p<.001$  to  $p=.001$ ). These were still under my significance threshold of  $p<.05$ , and so the overall results stand: H1 and H2 continued to be supported, while H3, H4, and H5 again found statistically significant difference in the opposite direction predicted. I believe this was a necessary first step in comparing analysis methods; if differences arise between the two, it does not appear that a difference in corpus structure would be at fault.

**Table 5**

*Means, Standard Deviations, and One-Way Analysis of Variance of Moral Content Using LIWC2015 and the Moral Foundations Dictionary 2.0 with Segmented Corpus.*

Moral Foundation	Biden		Trump		$F(1, 209)$	$p$
	$M$	$SD$	$M$	$SD$		
Care	.97	.48	.72	.33	19.57	<.001
Fairness	.36	.27	.23	.14	19.42	<.001
Loyalty	1.16	.45	.93	.52	9.04	.003
Authority	1.02	.41	.73	.23	43.01	<.001
Sanctity	.60	.27	.49	.22	10.88	.001

After completing this intermediate step, I analyzed the segmented corpus with eMFDScore using the Moral Foundations Dictionary 2.0 as the reference dictionary. While eMFDScore will not return absolute word counts as LIWC2015 does, it should depict a pattern of moral content that can be compared to the pairwise analysis performed above. If these agree, it again increases confidence in a comparison between methods. I describe this process in detail below.

### ***Sample***

For this analysis, I used the segmented corpus as described above.

### ***Design***

This analysis tested the moral content of the segmented corpus described above in a manner similar to experiment 1. The difference here lies in how the scores were calculated for each moral foundation. Rather than reporting keyword density, eMFDScore reports the number of words per moral foundation divided by the total number of words identified across all foundations. This gives relative scores per moral foundation without regard to overall word count. The purpose of this step is to compare results to the candidates' moral orientation patterns as observed in experiment 1.

### ***Procedure***

I used the eMFDScore program to analyze the segmented corpus using the Moral Foundations Dictionary 2.0 as a reference. I combined virtue/vice scores into a single score per category before loading the data into IBM SPSS 28 for ANOVA comparisons.

### ***Results***

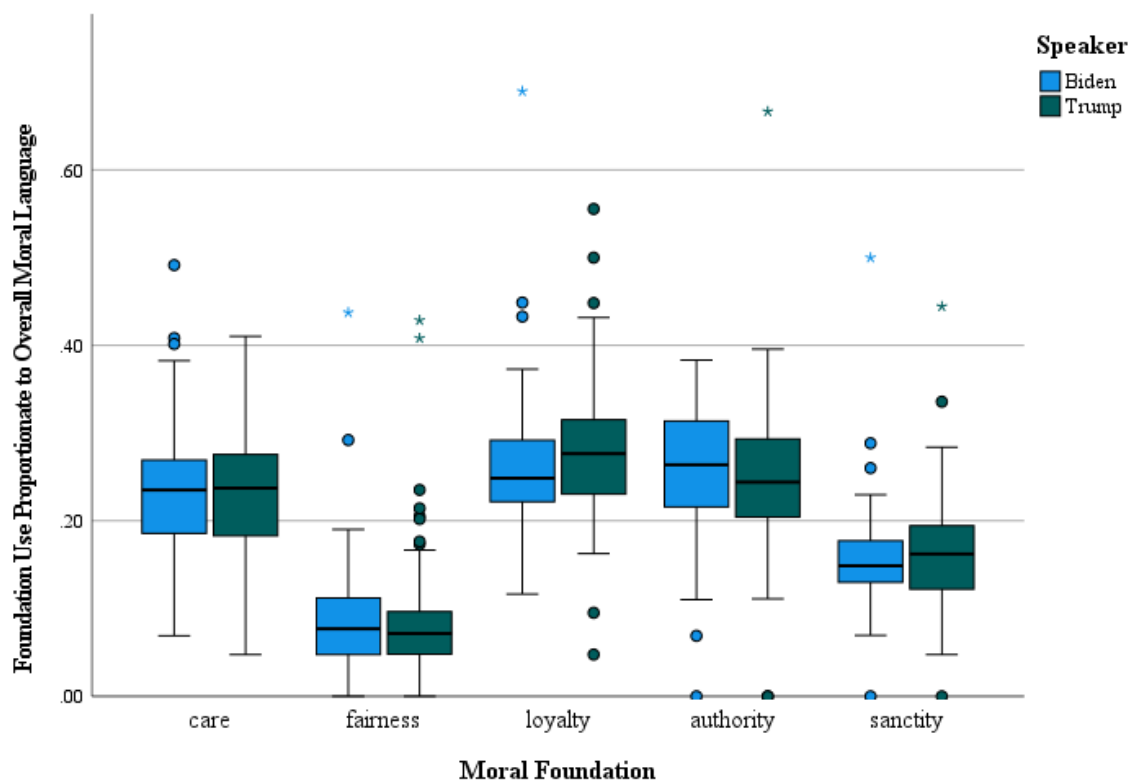
Rather than the keyword density scores that LIWC2015 provided, eMFDScore showed the keyword use per category divided by total moral keywords found. See figure



2 for a boxplot of this analysis, along with tables 6 and 7 for descriptive statistics and ANOVA results. The ANOVA indicated that the mean relative proportions of moral content from each foundation did not vary between candidates at a statistically significant level.

**Figure 2**

*Boxplot Depicting Moral Keyword Analysis Using eMFDScore and the Moral Foundations Dictionary 2.0.*



**Table 6**

*Descriptive Statistics of eMFDScore Analysis Using the Moral Foundations Dictionary 2.0 with Segmented Corpus.*

	Care	Fairness	Loyalty	Authority	Sanctity
Biden <sup>a</sup>					
Min	.07	.00	.12	.00	.00
Max	.49	.44	.69	.38	.50
M	.24	.09	.26	.26	.16
SD	.08	.07	.09	.08	.07
Skew	.67	2.86	2.26	-.77	2.53
Trump <sup>b</sup>					
Min	.05	.000	.05	.00	.00
Max	.41	.43	.56	.67	.44
M	.23	.08	.28	.25	.16
SD	.06	.06	.07	.07	.06
Skew	-.17	2.94	.49	.89	.86

<sup>a</sup>*n* = 59. <sup>b</sup>*n* = 152.

**Table 7**

*Means, Standard Deviations, and One-Way Analysis of Variance of Moral Content Using eMFDScore and the Moral Foundations Dictionary 2.0 with Segmented Corpus.*

Moral Foundation	Biden		Trump		<i>F</i> (1, 209)	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Care	.24	.08	.23	.06	.25	.615
Fairness	.09	.07	.08	.06	.56	.454
Loyalty	.26	.09	.28	.07	1.62	.204
Authority	.26	.08	.25	.07	.59	.445
Sanctity	.16	.07	.16	.06	.49	.487

## Discussion

These results did not support my hypotheses, but it is worth noting that this is due to a different operationalization of the moral content analysis. What was more noteworthy about these results is that they supported an observation from the first analysis using LIWC2015; while Biden used a higher keyword density across all moral

foundations, the overall pattern of moral content does not significantly vary between the two candidates.

By performing these intermediate analyses, I have addressed a gap left by Hopp et al. (2020) in comparing their method with previous methods. Using the LIWC2015 program alongside the Moral Foundations Dictionary 2.0, I have shown that the analysis of my sample did not significantly vary between an unmodified corpus versus a segmented corpus arranged for use with the eMFDScore program. Further, I demonstrated that the measurement, and thus operationalization, of moral content was fundamentally different between a LIWC2015 analysis and an eMFDScore analysis when using the same reference dictionary. While the eMFDScore analysis did confirm a pattern observed in the LIWC2015 analysis, the two sets of results lead to different conclusions when testing my hypotheses. With these results in mind, I turn now to an analysis using the eMFDScore program with the extended Moral Foundations Dictionary as a reference.

## Experiment 3

### Method

#### *Sample*

For this analysis, I used the segmented corpus described in experiment 2.

#### *Design*

For my final analysis, the independent variable was again the speaker, either Biden or Trump. The unit of analysis was an individual speech or speech segment, analyzed as a segment of text with 6,000 words or less. The dependent variables were the measure of moral content for each moral foundation, operationalized here as the mean eMFD annotation score of all identified words in each segment.

#### *Procedure*

I used the eMFDScore program to analyze the segmented corpus with the extended Moral Foundations Dictionary as the reference word list. I configured eMFDScore to return total probabilities across all moral foundations in order to capture the continuous moral weighting scores for which Hopp et al. (2020) advocated. eMFDScore also returns “sentiment” scores for each foundation, but since I combined virtue/vice scores in the previous analyses, I had no direct comparison for the sentiment scores. I loaded the moral foundation scores into IBM SPSS 28 for ANOVA comparisons.

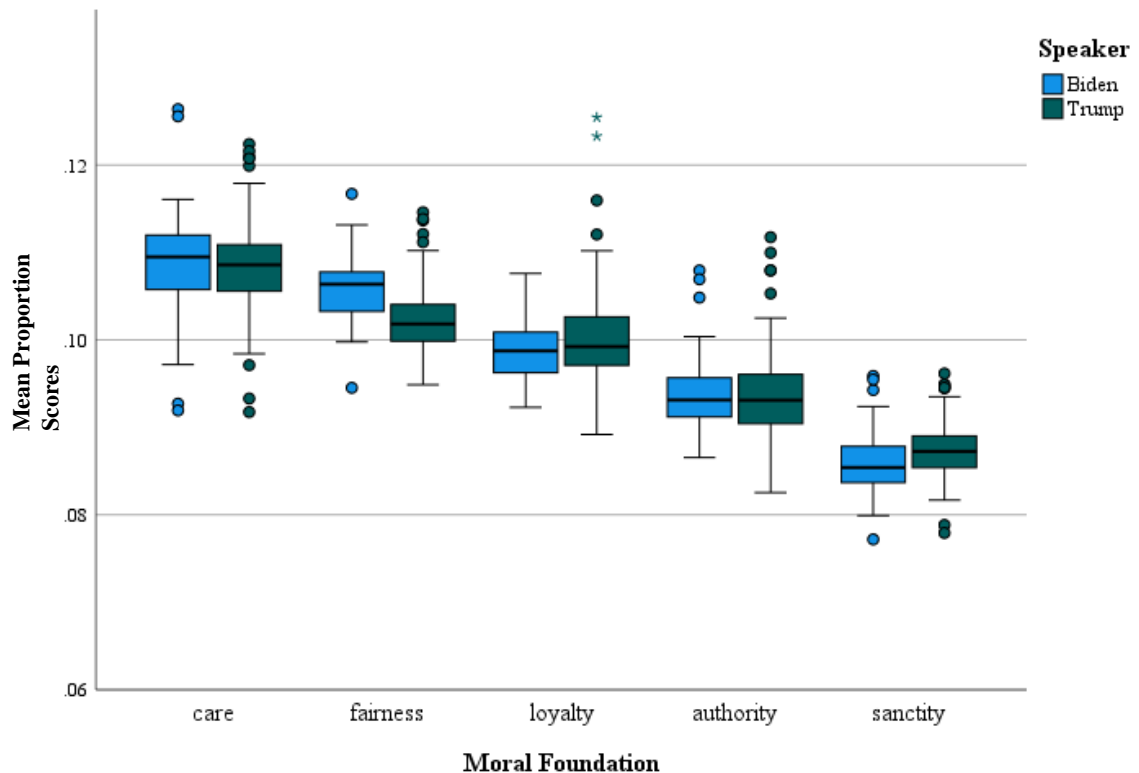
### Results

When using eMFDScore with the Moral Foundations Dictionary 2.0, I obtained scores that indicated keywords per moral foundation divided by total keyword use. However, the scores using the extended Moral Foundations Dictionary are calculated differently. Words in this dictionary are assigned multiple scores per foundation, with

each score indicating the proportion of initial participants that assigned a word to a particular foundation. For each corpus segment analyzed, the total word scores are added and then averaged within each foundation. A boxplot depicting these mean scores per foundation is shown in figure 3.

**Figure 3**

*Boxplot Depicting Moral Keyword Analysis Using eMFDScore and the extended Moral Foundations Dictionary.*



My initial observation of the data is that both candidates again appeared to exhibit the same pattern of moral content, but this pattern is very different from the one depicted by the previous analyses using the Moral Foundations Dictionary 2.0. See table 8 for descriptive statistics of these results, and table 9 for the ANOVA results.

**Table 8**

*Descriptive Statistics of eMFDScore Analysis Using the extended Moral Foundations Dictionary with Segmented Corpus.*

	Care	Fairness	Loyalty	Authority	Sanctity
Biden <sup>a</sup>					
M	.11	.11	.10	.09	.09
SD	.006	.004	.003	.004	.004
Skew	-.11	-.11	.30	1.38	.58
Trump <sup>b</sup>					
M	.11	.10	.10	.09	.09
SD	.005	.004	.005	.005	.003
Skew	-.03	.82	1.83	.85	.22

<sup>a</sup> $n = 59$ , <sup>b</sup> $n = 152$ .

**Table 9**

*Means, Standard Deviations, and One-Way Analysis of Variance of Moral Content Using eMFDScore and the extended Moral Foundations Dictionary with Segmented Corpus.*

Moral Foundation	Biden		Trump		$F(1, 209)$	$p$
	$M$	$SD$	$M$	$SD$		
Care	.11	.006	.11	.005	.11	.741
Fairness	.11	.004	.10	.004	39.79	<.001
Loyalty	.10	.003	.10	.005	3.10	.080
Authority	.09	.004	.09	.005	.25	.616
Sanctity	.09	.004	.09	.003	10.18	.002

Regarding the care foundation, Biden's moral content ( $M=.109$ ,  $SD=.006$ ) was not significantly different from Trump's moral content ( $M=.109$ ,  $SD=.005$ ),  $F(1, 209)=.110$ ,  $p=.741$ ; this did not support H1, thus it contradicted findings of my first analysis.

Regarding the fairness foundation, Biden's moral content ( $M=.106$ ,  $SD=.004$ ) was significantly higher than Trump's moral content ( $M=.102$ ,  $SD=.004$ ),  $F(1, 209)=39.789$ ,  $p<.001$ ; this supported H2 and was in line with the findings of my first analysis.

Regarding the loyalty foundation, Biden's moral content ( $M=.099$ ,  $SD=.003$ ) was lower than Trump's moral content ( $M=.100$ ,  $SD=.005$ ),  $F(1, 209)=3.104$ ,  $p=.08$ . This

result approached the significance level of  $p=.05$ , but it did not fully support H3, nor was it in line with the initial analysis.

Regarding the authority foundation, Biden's moral content ( $M=.094$ ,  $SD=.004$ ) was not significantly different from Trump's moral content ( $M=.094$ ,  $SD=.005$ ),  $F(1, 209)=.252$ ,  $p=.616$ . This did not support H4, nor was it in line with the initial analysis.

Regarding the sanctity foundation, Biden's moral content ( $M=.086$ ,  $SD=.004$ ) was significantly lower than Trump's moral content ( $M=.087$ ,  $SD=.003$ ),  $F(1, 209)=10.184$ ,  $p=.002$ . This supports H5, though this result contradicted the initial findings of the MFD 2.0 analysis.

To further explore these results, I ran another set of within subject comparisons (shown in table 10), as the boxplot appears to show a similar pattern of moral content use for each candidate, with the highest mean probability scores in the care foundation and descending scores for each subsequent foundation. Pairwise t-tests confirm this pattern, with both candidates showing statistically significant differences at the level  $p<.001$  for each step down the line.

**Table 10**

*Within-Subject Comparisons Using eMFDScore and the extended Moral Foundations Dictionary.*

	Biden			Trump		
	M	SD	<i>t</i> (58)	M	SD	<i>t</i> (151)
Pair 1						
Care	.11	.006	4.70***	.11	.005	22.83***
Fairness	.11	.004		.10	.004	
Pair 2						
Fairness	.11	.004	12.94***	.10	.004	6.95***
Loyalty	.10	.003		.10	.005	
Pair 3						
Loyalty	.10	.003	12.26***	.10	.005	32.04***
Authority	.09	.004		.09	.005	
Pair 4						
Authority	.09	.004	16.11***	.09	.005	20.86***
Sanctity	.09	.004		.09	.003	

\*\*\*  $p < .001$  (two-tailed).

### ***Correlations Between Dictionaries***

As one final comparison of the Moral Foundations Dictionary 2.0 and the extended Moral Foundations dictionary, I ran a correlation analysis between the results obtained by both dictionaries using the eMFDScore program. Hopp et al. (2020) performed the same test and noted significant positive correlations between the two different scores reported for each of the five moral foundations. Thus, they concluded that both dictionaries reliably measure the same content, with their extended Moral Foundations Dictionary providing a more valid measure of that content. However, the scores obtained for my sample do not support the same correlations.

Using the scores obtained from the Moral Foundations Dictionary 2.0 and the extended Moral Foundations Dictionary, both as assessed by the eMFDScore program, I found a statistically significant positive correlation between scores in the care foundation,



$r(209)=.49, p<.001$ ; the loyalty foundation,  $r(209)=.42, p<.001$ ; and the sanctity foundation,  $r(209)=.15, p=.033$ . However, I found a non-significant correlation between scores in the fairness foundation,  $r(209)=-.05, p=.435$ , and a statistically significant *negative* correlation between scores in the authority foundation,  $r(209)=-.22, p=.001$ .

## Discussion

This round of analysis revealed a different picture overall than any of the analyses using the Moral Foundations Dictionary 2.0. Here, both candidates appear to have emphasized the ‘care’ and ‘fairness’ foundations over the rest, which seems to reflect a liberal model of morality as described by Moral Foundations Theory. As with the first analyses, these results showed nearly identical patterns of moral foundation content between the two candidates. While two categories, ‘fairness’ and ‘sanctity,’ did contain statistically significant differences in alignment with my hypotheses, these differences are on quite a small scale. It is unclear then what substantive difference may truly exist between the two candidates under this analysis.

It is interesting that the two dictionaries did not significantly correlate across all five moral foundations as Hopp et al. (2020) found. My sample contained no significant correlation between the measure of fairness between the two dictionaries, thus I can only conclude that the two candidates used words that the extended Moral Foundations Dictionary coded under ‘fairness’ that were not included in the same category in the Moral Foundations Dictionary 2.0. The presence of a negative correlation for the ‘authority’ is perhaps even more striking, possibly indicating opposing bases for how the concept of authority appears in the corpus.

## General Discussion

### Theoretical Implications

When using an automated analysis to extract keywords from a text, it is inevitable that a researcher will confront the question of meaning. Is it truly possible to understand the moral meaning behind a text using these results? The underlying assumption of the LIWC2015 analysis is that yes, it is possible to gain some degree of understanding about the author or speaker's moral content. Frimer (2019) briefly describes the history of analyzing words to uncover the thoughts and intentions of the speaker of those words, a history also recounted in Tausczik and Pennebaker's (2010) description of how they developed Linguistic Inquiry and Word Count. However, Tausczik and Pennebaker (2010) offered a strong caveat about what their methods can and cannot accomplish, acknowledging that LIWC strips the meaning from words due to imprecise measurement and a lack of context (p. 30).

It seems almost necessary then to couple any attempt at an automated word count analysis with a qualitative study to determine the validity of the moral foundations highlighted by the analysis and arrive at a real understanding of the underlying meaning of this analysis. For example, does Biden's average moral keyword density of 1.19% in the loyalty foundation correspond to any particular message about loyalty, or is a researcher simply meant to brush aside this question and accept that this quantity of words is sufficient to reveal an underlying psychological inclination towards loyalty? The latter almost appears to be the case, as Tausczik and Pennebaker's (2010) caveats do not appear to be present in much of the research using LIWC and the Moral Foundations Dictionaries.

The results from experiment 3 show a pattern of morality for each candidate with statistically significant differences between various foundations, both within and between candidates. Yet the scores themselves fall within a very tight range, from .09 to .11, all with tight standard deviations of .003 to .006. These scores give a particularly tenuous foundation on which to draw substantive conclusions about the candidates' moral stances. They appear to point towards a conclusion that an audience might interpret a higher emphasis on the 'care' foundation, but only by a matter of two percentage points over the lowest foundation of 'sanctity.' Does this truly indicate anything substantial about the moral content in the corpus?

## **Practical Implications**

### ***Social Context***

Where all methods converge in my study is the result that both candidates used quite similar patterns of language related to the five moral foundations. While I began this study with a general question about the possibly polarized nature of Biden's and Trump's moral stances, there now appears to be little partisan difference between them as would be predicted by the relevant theory. I return to the research by Cienki (2005) and Deason and Gonzalez (2012). While these studies viewed morality through Lakoff's (1996) models, they did find similar evidence that presidential candidates do not fall neatly in line with the model. Deason and Gonzalez (2012) specifically point out the relevance of social context and dominant issue frames that may dictate, to some extent, the language that candidates will use to discuss an issue. This could very well be the case with the current study, and a full contextual analysis could warrant future research.

Briefly though, it is worth considering some of the societal context of the 2020 election and how this might have affected language use by the candidates.

On March 13, 2020, President Trump declared a national state of emergency concerning COVID-19, a newly emerged coronavirus pandemic. This virus upended social life and is yet to be fully contained as of this writing, a year and a half after Trump's declaration. Several issues relating to the virus became political touchstones, such as whether the government had the right to mandate that all citizens wear protective face masks, limit indoor gatherings, or receive vaccinations. This could very well have created an issue culture surrounding the topics of authority, loyalty, and even care, thus dictating certain ways of discussing these topics on the national stage. As the pandemic hit nations worldwide, so did issues of social and racial justice in response to the death of George Floyd, as previously mentioned. Many public entities, such as entertainment companies and even food brands, came to terms with elements of their business that perpetuated inequality in some form. This could have affected the way that people discussed fairness in the public sphere.

### ***Intent and Interpretation of Moral Messages***

It is common practice for politicians to hire professional speech writers, particularly when campaigning for office. These speeches are written deliberately and for a specific purpose. Perhaps then it is no wonder that both Biden and Trump appeared to have such similar moral orientations, as this may just reflect the language deliberately chosen to appeal to a wide audience. This may be part of the reason that Graham et al. (2009) felt that political speeches would be ineffective in validating the Moral Foundations Dictionary, and for a validation study they would appear to be correct.

However, the two methods of moral analysis presented here seem quite appropriate to study speeches when considering the two concepts of deliberate language use and audience appeal. After all, the Moral Foundations Dictionary 2.0 is inherently built upon the assumed and observed frequency of words pertaining to morality, while the extended Moral Foundations Dictionary is inherently built upon the interpretation of words pertaining to morality.

Both methods could be useful then, for such instances as deliberate speech writing. Before the speech is delivered, there would be an opportunity to analyze the speech with the Moral Foundations Dictionary 2.0 and validate that the intended moral messages were present. This could be followed by an extended Moral Foundations Dictionary to determine if an audience may be likely to interpret the message in the same way it was intended. This concept may underlie the differences between the results in my study, particularly for content in the ‘fairness’ foundation. These differences could point to a disconnect between the message that was intended and the message that might be interpreted.

In the current digital age, researchers have access to untold amounts of data, including transcribed speeches from practically every major politician in recent memory. It is tempting to assume that a simple, automated analysis would be able to reveal the content of a source text, neatly quantified and categorized into different categories of meaning. It is precisely this meaning, however, that gets lost when keywords are decontextualized. This is especially true of text that originated as speech, wherein the actual words spoken make up only a part of the intended and interpreted meanings involved.

## **Limitations and Conclusions**

There are at least three potential limitations to this study. First, I tested two individual speakers against a generalized model of moral language use. Perhaps, then, it should not be surprising that neither of them fit their predicted model. Graham et al. (2009) noted that they specifically avoided analysis of political speeches because of the assumption that the moral content would be scripted and altered to appeal to the voting public, rather than being indicative of the speaker's actual moral foundations. My results perhaps confirm this assumption, but future research could take a deeper look at this to determine the extent to which moral content analysis can reveal anything useful in political campaign speeches. It would also be useful to increase the number of speakers analyzed from each party, perhaps using a corpus of campaign speeches over time. Using a greater number of speakers across different social contexts could be a better test of the claims of Moral Foundations Theory. As it stands, my results should not be taken as a direct refutation of the underlying theory just because the patterns predicted by the theory did not hold true.

Second, my tests used their respective analysis methods at their most basic level. For example, I used combined moral foundation scores when using LIWC2015 rather than separately analyzing each candidate's vice/virtue split. Similarly, I did not use the moral sentiment ratings returned by eMFDScore, nor did I delve into advanced features of eMFDScore such as syntactic parsing which highlights moral agents and targets of a speech. It is possible that a more detailed look at the full array of possible data would yield yet another perspective on the moral patterns in my sample. However, it is also possible that this level of detail would obscure the picture even further. It would be useful

for future research to clearly quantify the differences between all available analysis methods, in their original and modified forms, to understand if one method is truly superior.

Finally, I did not include any form of qualitative analysis that might reveal a broader contextual understanding of the moral language of either candidate. Admittedly, this is a potential limitation of any automated keyword analysis. Computer programs that add up instances of words cannot identify what those words mean in context. However, the ability of computers is always advancing, and the syntactic parsing ability of eMFDScore (Hopp et al, 2020) is one possible step towards bridging the gap between computer and human analysis. As this and similar methods arise, it is essential that the developers of these methods demonstrate sound validity and reliability of their measures and a logical operationalization of the concepts under analysis.

This study presented several perspectives on the moral content of Biden's and Trump's campaign speeches during the 2020 US Presidential Election. Using two different reference word lists, it appeared that neither candidate used moral content in patterns predicted by Moral Foundations Theory. It does appear that Biden's speeches contained on overall higher density of moral content than Trump's, in line with his claim to be the morally grounded option of the two. More interestingly though, the patterns of moral content were very similar between both candidates, perhaps because of overarching public issues that dictated certain moral stances. This study adds to the research by identifying gaps in the comparison of two methods for moral content analysis, as well as identifying how both methods may be used to present a more detailed perspective of this analysis.

## References

- Baker, P. (2017, September 24). A divider, not a uniter, Trump widens the breach. *The New York Times*. <https://www.nytimes.com/2017/09/24/us/politics/trump-divisiveness.html>
- Baker, P. (2020, September 17). For Trump, it's not the United States, it's red and blue states. *The New York Times*. <https://www.nytimes.com/2020/09/17/us/politics/trump-america.html>
- Belson, K., & Davis, J. H. (2017, September 23). Trump attacks Warriors' Curry. LeBron James's retort: 'U Bum.' *The New York Times*.  
<https://www.nytimes.com/2017/09/23/sports/football/trump-nfl-kaepernick.html>
- Biden, J. (2020, August 31). *Joe Biden Pittsburgh Speech Transcript August 31* [Transcript database]. Rev.Com. <https://www.rev.com/blog/transcripts/joe-biden-pittsburgh-speech-transcript-august-31>
- Deason, G., & Gonzales, M. H. (2012). Moral politics in the 2008 presidential convention acceptance speeches. *Basic and Applied Social Psychology*, 34(3), 254–268.  
<https://doi.org/10.1080/01973533.2012.674450>
- Dias, E. (2020, October 18). Biden and Trump say they're fighting for America's 'soul.' What does that mean? *The New York Times*.  
<https://www.nytimes.com/2020/10/17/us/biden-trump-soul-nation-country.html>
- Diaz, D. (2016, October 9). *3 times Donald Trump defended his "locker room" talk* | *CNN Politics*. CNN.Com. <https://www.cnn.com/2016/10/09/politics/donald-trump-locker-room-talk-presidential-debate-2016-election/index.html>



- Fact check: Clarifying the comparison between popular vote and counties won in the 2020 election. (2020, December 29). *Reuters*. <https://www.reuters.com/article/uk-factcheck-votes-counties-election-idUSKBN2931UY>
- Feinberg, M., & Willer, R. (2015). From gulf to bridge: When do moral arguments facilitate political influence? *Personality and Social Psychology Bulletin*, 41(12), 1665–1681. <https://doi.org/10.1177/0146167215607842>
- Frimer, J., Haidt, J., Graham, J., Dehghani, M., & Boghrati, R. (2017). *Moral foundations dictionaries for linguistic analyses, 2.0*. Unpublished Manuscript. Retrieved from: [www.jeremyfrimer.com/uploads/2/1/2/7/21278832/summary.pdf](http://www.jeremyfrimer.com/uploads/2/1/2/7/21278832/summary.pdf)
- Frimer, J. A. (2019). 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>
- Fury as Trump mocks Muslim soldier's mother Ghazala Khan. (2016, July 31). *BBC News*. <https://www.bbc.com/news/election-us-2016-36935175>
- 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*.
- 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>
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. In P. Devine &

- A. Plant (Eds.), *Advances in Experimental Social Psychology* (Vol. 47, pp. 55–130). Academic Press. <https://doi.org/10.1016/B978-0-12-407236-7.00002-4>
- Gstalter, M. (2019, March 2). *Trump mocks Sessions, mimics Southern accent during CPAC speech*. TheHill. <https://thehill.com/homenews/administration/432325-trump-mocks-sessions-southern-accent-during-cpac-speech>
- Haidt, J. (2013). *The righteous mind: Why good people are divided by politics and religion* (First Vintage Books edition.). Vintage Books.
- Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, 133(4), 55–66.  
<https://doi.org/10.1162/0011526042365555>
- Hopp, F. R., Fisher, J. T., Cornell, D., Huskey, R., & Weber, R. (2020). The extended Moral Foundations Dictionary (eMFD): Development and applications of a crowd-sourced approach to extracting moral intuitions from text. *Behavior Research Methods* 53, 232-246. <https://doi.org/10.3758/s13428-020-01433-0>
- Hopp, F. R., Fisher, J. T., & Weber, R. (2021). *eMFDScore* [Jupyter Notebook, Python]. Media Neuroscience Lab. <https://github.com/medianeuroscience/emfdscore>  
(Original work published 2019)
- Joe Biden. (2019, April 25). *Joe Biden For President: America Is An Idea*.  
<https://www.youtube.com/watch?v=VbOU2fTg6cI>
- Naylor, B. (2021, February 10). Read Trump’s Jan. 6 speech, a key part of impeachment trial. *NPR*. <https://www.npr.org/2021/02/10/966396848/read-trumps-jan-6-speech-a-key-part-of-impeachment-trial>

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2015). Linguistic inquiry and word count: LIWC2015 [Computer software]. *Austin, TX: liwc.net*.

Rakich, N. (2021, January 8). Trump helped take extremist views from the fringes of society to a mob attacking the Capitol. *FiveThirtyEight*.

<https://fivethirtyeight.com/features/trump-helped-take-extremist-views-from-the-fringes-of-society-to-a-mob-attacking-the-capitol/>

Skitka, L. J. (2010). The psychology of moral conviction. *Social and Personality*

*Psychology Compass*, 4(4), 267–281. <https://doi.org/10.1111/j.1751-9004.2010.00254.x>

Skitka, L. J., & Morgan, G. S. (2014). The social and political implications of moral conviction. *Political Psychology*, 35, 95–110.

Stohr, K. (2017, January 23). Our new age of contempt. *The New York Times*.

<https://www.nytimes.com/2017/01/23/opinion/our-new-age-of-contempt.html>

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words:

LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24–54. <https://doi.org/10.1177/0261927X09351676>

Taylor, J. (2016, October 7). “You can do anything”: In 2005 tape, Trump brags about

groping, kissing women. *NPR*. <https://www.npr.org/2016/10/07/497087141/donald-trump-caught-on-tape-making-vulgar-remarks-about-women>