Moral Foundations of Political Reasoning[*]
Investigating the Moral Underpinnings of Political Judgment

Patrick W. Kraft[†]

September 7, 2016

Abstract

Moral Foundations Theory (MFT) proposes that political attitudes and ideology are shaped and structured by fundamental moral intuitions. However, recent research cast some doubt on the theory's basic tenets and argued that moral foundations might be less stable than previously thought. The present paper directly explores the conditions under which moral foundations are expressed in politics. Using open-ended survey responses in the 2012 American National Election Study, I utilize a moral dictionary proposed in the MFT literature to identify references to basic moral intuitions when individuals report their attitudes towards political parties and candidates. The results show that the reliance on moral foundations in political reasoning is conditional on campaign exposure and political engagement, which provides a potential explanation for the instability of moral foundations in the realm of politics. The findings further indicate that moral foundations are evoked to justify beliefs and preferences in a given political context.

Keywords: Moral Foundations Theory, Political Reasoning, Ideology, Political Behavior, Open-ended Survey Responses

1 Introduction

To what extent are political belief systems and ideologies meaningfully structured by individual psychological characteristics and underlying motivations? This question has been of frequent scholarly interest in political science and related disciplines, yielding an array of different perspectives. Early accounts emphasized how ordinary citizens lack consistent political attitudes and knowledge necessary to form meaningful ideologies (e.g. Converse, 1964). However, with increasing levels of polarization and partisan sorting in contemporary politics (Iyengar and Westwood, 2015), there has been renewed interest in systematic psychological and attitudinal differences between liberals and conservatives (Jost, 2006). One such area of research focuses on the differing moral concerns of liberals and conservatives. According to *Moral Foundations Theory* (MFT), moral thinking is organized by five innate intuitions: harm/care, fairness/reciprocity, ingroup/loyalty, authority/respect, and purity/sanctity (Haidt and Joseph, 2008; Graham et al., 2011). Liberals and conservatives differ in their relative emphasis of these foundations, with liberals prioritizing the foundations of harm/care and fairness/reciprocity, and conservatives endorsing all five dimensions more or less equally (Graham, Haidt, and Nosek, 2009).

However, recent research cast some doubt on the theories basic tenets. Smith et al. (2016) empirically investigated three implicit assumptions underlying MFT, namely that moral foundations can be described as individual traits that are relatively stable over time, that changes in moral foundations predict changes in political ideology, and that the foundations are heritable. The empirical evidence provided by Smith et al. (2016) does not support any of these assumptions (but see Haidt's response). Other researchers voiced similar criticisms regarding MFT. For example, Suhler and Churchland (2011) argued that the definition innateness used by Haidt and colleagues is too loose and ultimately not backed by the empirical evidence put forward to support the theory. Instead, the authors discuss recent evidence from neuroscience and genetics that contradicts the theoretical arguments underlying MFT (Suhler and Churchland, 2011).

This critique of MFT stands in contrast to a series of studies that documented the political relevance of moral foundations. For example, a growing body of work provided evidence for the influence of moral foundations on vote choice (Iyer et al., 2010; Franks and Scherr, 2015), issue preferences (Koleva et al., 2012; Low and Wui, 2015; Clifford, Jerit, Rainey, and Motyl, 2015), and candidate trait evaluations (Clifford, 2014). As such, it does not seem to be time to abolish MFT and its taxonomy of moral considerations entirely. Rather, further research needs to directly examine the *conditions* under which moral foundations are consequential in the realm of politics. Indeed, Smith et al. (2016) conclude that their findings are "consistent

with the notion of moral foundations being highly responsive to context: more of a state than a trait". However, only very few studies systematically investigated the conditionality of the relationship between moral foundations and political attitudes (see Clifford, Jerit, Rainey, and Motyl, 2015, for a notable exception).

The present study provides a first step in this direction by directly analyzing the degree to which individuals vary in the expression of moral concerns in political reasoning and judgment. Using data from the 2012 American National Election Study, I create a measure of moral reasoning based on individual responses to the open-ended likes-dislikes questions and the moral dictionary introduced by Graham, Haidt, and Nosek (2009). Relying on open-ended questions rather than traditional survey-based measures of moral foundations allows me to investigate patterns of moral reasoning among individuals where the potential connection between morality and politics is not induced or facilitated by design.

The analyses reveal similar systematic differences between liberals and conservatives in their reliance on specific moral considerations when evaluating political parties and candidates even without being cued to think about morality. These differences, in turn, affect political evaluations as well as voting behavior—even after controlling for a persons party identification. Most importantly, the overall reliance on moral foundations in political judgment as well as the respective differences between liberals and conservatives are (at least partly) conditional on campaign exposure and political engagement. Even though the ideological differences are broadly consistent with MFT, moral reasoning in the realm of politics is more context-specific than previous research suggests. Additional to the substantive conclusions, the present study has important methodological implications since it illustrates the value of open-ended survey measures to illuminate the antecedents of political reasoning.

2   Theoretical Framework

Moral values serve as a source for coherence in political attitudes and they shape individual belief systems. There is evidence, for example, that moral considerations predict a person's attitudes on a range of "culture-war" issues such as abortion and same-sex marriage above and beyond demographic characteristics and ideology (Koleva et al. 2012; also see Clifford, Jerit, Rainey, and Motyl 2015). Even issues that are not considered intrinsically "moral," such as those related to the economy, can be connected to underlying moral convictions (Ryan, 2014). Moral convictions, in turn, have been shown to reduce tolerance of divergent attitudes and lower acceptance of compromise (Skitka, 2010; Ryan, 2016). As such, individual moral reasoning may provide important insights into the recent increase in polarization in American Politics (Iyengar and Westwood, 2015).

2.1   Moral Foundations Theory

Rather than being formed through explicit reasoning, Haidt (2001) argues that moral judgment is based on automatic and affective intuitions, which are in turn influenced by social and environmental factors. More specifically, moral intuition "appears to be the automatic output of an underlying, largely unconscious set of interlinked moral concepts. These concepts may have some innate basis [...], which is then built up largely by metaphorical extensions from physical experience" (Haidt, 2001, 825). According to this view, explicit moral reasoning is better described as a rationalization of these intuitions, which are partially innate and therefore "organized, to some extent, in advance of experience" (Haidt and Joseph 2008, 367, but see Suhler and Churchland 2011). Haidt and Joseph (2008) identified five basic intuitions which build the psychological foundations of human morality and are inherently linked to evolutionary adaptive challenges that appear cross-culturally.

One major focus of the moral foundations framework is the relationship between moral values and political belief systems (c.f. Haidt, 2012). Liberal morals have been shown to focus on "individualizing" foundations, which include harm/care and fairness/reciprocity. Conservatives, on the other hand, also emphasize the remaining foundations of ingroup/loyalty, authority/respect, and purity/sanctity, labeled the "binding" foundations (Haidt and Graham, 2007; Graham, Haidt, and Nosek, 2009).[1] In order to measure individual differences in the emphasis of each foundation, Haidt and colleagues developed the Moral Foundations Questionnaire (MFQ). The MFQ consists of a series of items that explicitly ask people to rate the relevance of different considerations when making decisions about right and wrong.[2] The questionnaire also asks individuals to indicate their level of agreement with statements that represent the values implied by the five foundations.[3]

Relying on the MFQ, subsequent research extended the initial findings. For example, scholars showed that moral concerns predicted attitudes towards a wide variety of divisive political issues (e.g. Koleva et al., 2012; Low and Wui, 2015). Federico et al. (2013) linked moral foundations to individual social dominance orientation (SDO) and right-wing authoritarianism (RWA). Further research directly investigated the relationship between moral foundations and candidate preferences (Iyer et al., 2010) or trait inferences about candidates (Clifford, 2014). Moral foundations have also been shown to predict turnout (Johnson et al.,

---

[1]Subsequent accounts of MFT discussed the inclusion of further dimensions, such as *Liberty/Oppression* (c.f. Graham et al., 2013; Haidt, 2012). However, the analyses presented here will only focus on the dimensions initially suggested in Haidt and Joseph (2008).

[2]For example, one of the considerations is "Whether or not some people were treated differently than others." High relevance of this statement is viewed das an indicator for the fairness/reciprocity dimension (c.f. http://www.moralfoundations.org/).

[3]For example, respondents are asked to report their agreement with the statement "I am proud of my country's history" as an indicator for the ingroup/loyalty dimension.

2014) as well as voting behavior in the 2012 US Presidential election (Franks and Scherr, 2015). Overall, this research strongly supports the view that liberals and conservatives endorse different moral foundations and that these differences are related to political attitudes, evaluations, and behavior.

Smith et al. (2016), on the other hand, used the MFQ to examine some basic assumptions underlying MFT using different twin samples. More explicitly, the authors investigated whether moral foundations can be viewed as stable traits that vary little over time, whether changes in moral foundations predict changes in ideology (or vice versa), and whether there are sufficient signs of heritability in the emphasis of moral foundations. Overall, the evidence reported in Smith et al. (2016) paints a gloomy picture for MFT: the foundations are less stable than ideology, do not seem to predict changes in the latter over time, and do not seem to be inherited through genes. However, it should also be noted that the analyses presented by Smith et al. (2016) are not without issues. For example, their analyses of stability over time relied on different sub-scales of the MFQ in each of the waves, which might induce some of the variability in the sample. Nevertheless, the results point towards the conclusion that the relationship between ideology and moral foundations might not be as stable as initial research suggested.

As such, directly investigating the conditions under which moral reasoning occurs when people think about politics might shed some light on the divide between the growing body of research that highlights the importance of MFT for political attitudes on the one hand, and the results presented by Smith et al. (2016) on the other hand. As I will argue below, this step involves measuring moral foundations in political reasoning without relying on the MFQ, which has been used in nearly all existing work (but see Clifford, 2014).

2.2   The 'Missing Link' to Political Reasoning

While many studies report consistent differences between liberals and conservatives in terms of their latent emphasis on moral foundations, recent research points to the instability of moral foundations and their relationship with ideology. In an attempt to bridge this divide, I focus on the question whether people utilize the foundations in their day-to-day political reasoning (i.e., without being prompted by the language of a questionnaire. Previous research has relied almost exclusively on the Moral Foundations Questionnaire (MFQ), which explicitly asks people to judge the relevance of particular moral considerations (e.g. Graham et al., 2011). However, by directly asking people about the importance of considerations related to the five foundations, previous analyses presuppose an important link that requires more careful empirical investigation. The present study addresses this gap by examining whether

differences in moral reasoning between liberals and conservatives manifest themselves in a more unobtrusive context (i.e., without explicitly asking people to think about morality).

Measuring moral foundations using the MFQ imposes important constraints on our theoretical conceptualization of moral reasoning and the range of hypotheses we can investigate. Most importantly, it does not allow us to directly examine the conditions under which the connection between moral values and political ideology manifests itself when citizens reason about politics and evaluate political actors. Some scholars have criticized the MFQ because it does not ask people to make moral judgments per se. Indeed, Graham, Haidt, and Nosek (2009, 1031) describe the reports on moral relevance as "self-theories of moral judgment," rather than direct measures of judgment itself. Such abstract self-theories might, in turn, deviate from actual judgments in specific situations (see Clifford, Iyengar, Cabeza, and Sinnott-Armstrong, 2015, for an alternative way to measure moral judgment).

From a theoretical perspective, moral foundations are viewed as stable predispositions that affect attitudes and preferences regardless of potential individual differences and contextual effects. However, the incorporation of moral considerations in a specific political context could be much more variable. Previous research suggests that campaigns and elite communication can have important influences on individual moral reasoning. For example, Clifford and Jerit (2013) found that at the elite level, proponents and opponents of stem cell research place distinctive weights on moral foundations which in turn affected the public attitudes and the underlying considerations related to the issue. A subsequent study showed that elite rhetoric plays an important role in linking individual moral foundations with political attitudes, but only for those individuals who were most likely be exposed to it (Clifford, Jerit, Rainey, and Motyl 2015; also see Day et al. 2014). While these studies do not contradict MFT, they cast some doubt on the notion that moral reasoning in politics is as a direct reflection of stable moral intuitions.

The literature on *moral convictions* further suggests that moral reasoning in the realm of politics can be more variable than proposed by MFT. Skitka and colleagues argue that individuals hold moralized attitudes if they have a universal perception of "right and wrong" connected to the issue at hand (Skitka, Bauman, and Sargis, 2005; Mullen and Skitka, 2006; Skitka, 2010). This view implies that there can be considerable variance in the degree to which moral considerations are raised between individuals and across issues.

Research on moral conviction and moral foundations developed largely independent of each other, which can be partly explained by the fact that the measurement approaches in both fields are not directly compatible. Moral foundations are measured as general predispositions that are implicitly assumed to apply in any context, whereas convictions are intrinsically linked to an individual's beliefs about specific issues. Moving beyond the MFQ

and measuring moral reasoning as the explicit expression of foundations can help us to bridge the gap between both literatures. If we assume that moral convictions can be expressed as moral considerations that overlap with the moral taxonomy developed in the Moral Foundations literature, then the variability in moral reasoning measured through open-ended responses can be informative for the degree to which political attitudes are infused by moral convictions.

Overall, individuals who are exposed to the political process and resulting elite communications are expected to be more likely to incorporate moral considerations in their political reasoning because they adopt the respective arguments from elite discourse. Individuals who are not engaged in politics, on the other hand, might focus on other considerations when thinking about and forming their political preferences. The extent to which individuals rely on moral considerations when evaluating political actors is therefore not necessarily stable among individuals and across contexts. Rather, the tendency to emphasize moral foundations may be contingent upon individual levels of political sophistication, media exposure, and political discussions.[4] The degree to which political attitudes are moralized as well as the ideological differences in moral reasoning could not be investigated by relying solely on MFQ. Instead, it is necessary to differentiate between a person's tendency to endorse a moral *foundation* (as stable predispositions), and moral *reasoning* as the actual reliance on specific moral considerations and arguments, both from a theoretical as well as a measurement perspective. Previous research in MFT largely neglected this difference.

3 Empirical Analyses

3.1 Overview and Hypotheses

Insofar as moral intuitions play a role in political reasoning, citizens should rely on the moral foundations when reporting their attitudes towards political actors, even if they are not explicitly asked to do so. Thus, the first step of the analyses will be to replicate the findings connected to moral foundations and ideology using open-ended survey responses:

---

[4]However, some studies indicate that individuals make frequent references to values when discussing their policy preferences (Feldman and Zaller, 1992) and that the reliance on basic values is not contingent on individual characteristics such as political sophistication (e.g. Goren, 2001, 2004; Marietta and Barker, 2007).

Hypothesis 1: Liberals will be more likely to spontaneously mention the moral foundations of harm/care and fairness/reciprocity than conservatives when evaluating political parties and candidates. Conversely, conservatives will be more likely to emphasize moral foundations of ingroup/loyalty, authority/respect, and purity/sanctity than liberals.

Moral reasoning measured through open-ended responses is further expected to be politically consequential. First of all, the differences in the degree to which each foundation is emphasized should directly affect political preferences and vote choice:

Hypothesis 2: References to individual moral foundations predict candidate and party preferences, as well as vote choice consistent with the propositions of MFT.

After replicating the basic finding in the MFT literature in a more unobtrusive survey context that does not evoke issues of morality (H1) and showing that these differences have substantial effects on subsequent political outcomes (H2), the present study contributes to the literature by clarifying the role of the foundations in day-to-day reasoning. Rather than being viewed as stable predispositions, individuals are expected to differ in the degree to which they rely on moral considerations in the realm of politics. I therefore argue that the use of the moral foundations is more variable and context specific than previous research on MFT suggests; more specifically:

Hypothesis 3a: Individuals who are more exposed to political discourse and more engaged in politics will be more likely to emphasize moral foundations when evaluating political parties and candidates than those who have less experience and are less politically engaged.

Hypothesis 3b: Ideological differences in the emphasis on moral foundations will be more pronounced for individuals who are more exposed to political discourse and more engaged politics than those who have less experience and are less politically engaged.

3.2   Data, Variables, and Model Specification

The analyses presented here are based on the 2012 American National Election Study, which contains two representative cross-sectional samples. One sample was conducted by computer assisted face-to-face interviews while the other sample is based on an internet panel group.

Both samples are pooled in the analyses. While each consisted of a pre-election and a post-election wave, most items described below are drawn from the pre-election wave.[5]

The major dependent variables are based on open-ended questions in which respondents were asked to report what they *liked* and *disliked* about either presidential candidate as well as the Republican and Democratic parties. More specifically, respondents were asked to list anything in particular that they like/dislike about the Democratic/Republican party as well as anything that might make them vote/not vote for either of the Presidential candidates and were probed by the interviewer asking "anything else?" until the respondent answered no. All open-ended responses were pre-processed by correcting spelling errors using an implementation of the Aspell spell checking algorithm in R (`www.aspell.net`), and deleting individuals who responded in Spanish. All responses to the eight open-ended like/dislike questions (evaluating both parties and both candidates) were aggregated for each individual.

Our goal is to use the open-ended responses to identify the extent to which respondents relied on specific moral foundations when talking about their political attitudes and preferences. One of the studies by Graham, Haidt, and Nosek (2009) consisted of a quantitative analysis of sermons from liberal and conservative churches. The authors proposed a dictionary of words (and word stems) that signal references to the specific moral foundations and showed that liberal sermons were more likely to contain expressions related to harm/care and fairness/reciprocity, than conservatives (see also Haidt, Graham, and Joseph, 2009, for multi-dimensional conceptualizations of ideology). Other studies in the domain of Moral Foundations Theory relied on a similar dictionary to identify references to moral dimensions in news media coverage about stem cell research (Clifford and Jerit, 2013).[6] In the present research, I use the same dictionary to analyze open-ended survey responses. To give an intuition about the content of the dictionary, consider the following examples. Words like "protect" and "suffer" indicate references to the harm/care foundation, "equality" and "tolerant" signal reasoning based on the fairness/reciprocity foundation, "patriot" and "betrayal" indicate reference to the ingroup/loyalty foundation, "honor" and "respect" signal considerations related to authority/respect, whereas "integrity" and "duty" indicate reference to the purity/sanctity foundation. Conventional dictionary-based methods consist of dichotomous indicators, simple counts, or proportions of signal word occurrences for each dictionary category in each text or response.

But what if the content of the dictionary itself is not perfectly suited for our purposes? While Graham, Haidt, and Nosek (2009) developed the word lists as a collection of general

---

[5]The open-ended items were included only in the pre-election wave. Accordingly, wherever possible, the set of explanatory variables was limited to the pre-election wave.

[6]The dictionary provided by Graham, Haidt, and Nosek (2009) is also presented in Appendix B.

associations related to the base foundations that could be applied in any context, some words included in the lists might be problematic when applied to the open-ended responses considered here. More specifically, some words are too ubiquitous in the realm of politics and candidate evaluations to be regarded as an unambiguous indicator for specific moral considerations. For example, the moral foundations dictionary includes "leader", "respect", "law", and "illegal" as a signal words for the authority/respect dimension. If respondents are asked about their attitudes towards both presidential candidates, they might be inclined to describe them as good or bad *leaders* irrespective of their endorsement of particular moral foundations.

One way to address this problem would be to revise the dictionary. However, this leaves a lot of discretion to the researcher and creates the potential issue that signal words are chosen based on ex-post reasoning about individual response behavior in the given context. Instead, I rely on techniques developed in the fields of information retrieval to match search queries to relevant text documents (see Manning et al., 2008, for an introduction). Consider the example of a political scientist who searches for articles containing the keywords "morality" and "politics" on Google Scholar. Should the search engine consider both terms as equally important when searching for matches in the text documents? If we were to simply sum up the number of matches for each document, we would end up identifying a lot of documents that are, in fact, too general (i.e. that only discuss "politics"). In order to identify the set of most relevant documents, it makes sense to increase the importance of keywords that appear in fewer documents (i.e. "morality"), since those terms carry more discriminative information.

I rely on the same logic to create a measure that represents the degree to which a person's open-ended response resembles Graham et al.'s (2009) dictionaries for each moral dimension (referred to below as "similarity score"). Formally, similarity is the tf-idf cosine score[7] between the dictionaries and a person's response to the open-ended items on the ANES. A technical discussion appears in Appendix C, but the measure has several useful properties that I note here. Most importantly, words that appear in nearly all open-ended remarks affect similarity less than the words mentioned only by few respondents (because ubiquitous words convey less information about differences across individuals). A related property of the measure is that words that are mentioned multiple times in a response are weighted more than words that only appear once. Similarity is a continuous measure that ranges between 0 (no overlap between dictionary and response) and 1 (response is identical to dictionary),

---

[7]The acronym tf-idf stands for "term frequency - inverse document frequency," which refers to the rationale that the frequency of specific terms are weighted by the inverse of the frequency of occurrence across documents.

and is independent of response length.

Overall, the similarity score provides a correction for potential distortions due to sub-optimal terms in the dictionary while leaving its exact content outside of the researcher's discretion. While the resulting measure theoretically ranges from 0 to 1, we only observe very low values since open-ended responses are naturally quite distinct from the raw dictionaries. Since the exact values do not have a straightforward interpretation[8], the measure for each moral dimension is rescaled to unit variance.

Respondents were not included in the analysis if they failed to provide an answer to all open-ended items, or if the interview language was Spanish. Table B.1 in the appendix provides an overview over the number of omitted cases. About 4% of the interviews were held in Spanish and about 7% of the respondents did not provide any open-ended response. Furthermore, Figure B.1 in the appendix displays histograms of the length of the respondents' answers to all open-ended items. On average, the collection of all open-ended responses consists of about 75 words for each individual.

Since the similarity scores are bound at zero (i.e. none of the words in the dictionary appear in the response), individual response patterns are modeled via Tobit regressions for each of the moral foundations under consideration. As will be further described below, the Tobit framework allows us to decompose the estimates into the effect on the probability of mentioning a specific foundation *at all* as well as the degree of emphasis on the foundation given that it was mentioned by a respondent (see McDonald and Moffitt, 1980, for details on the decomposition of Tobit model estimates). The key independent variable used to predict the emphasis on each of the moral foundations, is *political ideology*. Respondents were asked to place themselves on a seven-point scale ranging from extremely liberal to extremely conservative. Since there is no reason to suggest that moderates should fall in between liberals and conservatives in terms of their moral foundations (i.e. that the relationship between 'continuous' ideological self-placement and the likelihood to mention specific moral foundations is inherently linear), I constructed dichotomous variables indicating whether respondent identified as liberals, conservatives, or moderates.

Additional control variables included in the analyses are *church attendance*, *education* (college degree), *age*, *sex*, *race* (African American), survey mode (online vs. offline), as well as the overall length of the individual responses in the open-ended questions (*measured as logged number of words*). The inclusion of the length of individual responses as control variables should account for potential confounding factors such as general effects of increased

---

[8]Technically, values on this variable describe the cosine of the angle between the tf-idf vector space representation of the dictionary and each open-ended response. Please refer to the Appendix C for a more detailed discussion.

political literacy on the complexity of open-ended responses.

In order to examine the relevance and consequences of moral reasoning, the similarity scores for each of the moral foundations are used as independent variables to predict political outcomes. The dependent variables considered here are *candidate* and *party evaluations* (measured as the respective feeling thermometer differentials), as well as *voting behavior* (measured as a dichotomous indicator of vote choice for the Democratic rather than the Republican Presidential candidate). In addition to the controls discussed previously, these analyses include measures of *party identification*.

The final set of analyses investigates whether the expression of moral considerations in political judgment is conditional on campaign exposure and engagement. The factors that are therefore expected to be related to references to moral foundations include *political sophistication*, which was measured as the sum of correct answers to objective knowledge questions. The analysis also investigates the effect of *political media exposure*, the frequency of *political discussions* with friends and family members, *turnout* in the previous election, as well as other forms of *non-conventional participation* (participation in demonstrations, wearing campaign buttons, or signing petitions). Here, we not only look at the influence on individual foundations, but also examine whether these factors influence *general* moral reasoning. This latter variable is measured as the sum of individual similarity scores, which can be interpreted as a aggregate measure of how much respondents emphasize any moral consideration in their responses.

3.3   Results

3.3.1   Ideological Differences in Moral Reasoning

Figure 1 provides an overview of the response patterns for individuals who identified as liberals, conservatives, or moderates. For each group, the figure displays the (weighted) proportion of respondents who mentioned words that were included in the five different moral foundations dictionaries as well as their 95% confidence intervals. Since responses for each individual represent their likes and dislikes across all eight open-ended items, each proportion indicates the percentage of individuals who mentioned a signal word belonging to the respective moral foundation in any of his or her open-ended responses evaluating the parties or candidates.

The patterns are largely consistent with theoretical expectations. Liberals were more likely than conservatives or moderates to mention the harm/care foundation. Almost half of the respondents identifying as liberals mentioned words belonging to the harm/care category in their responses. Furthermore, they were more likely than conservatives to mention the
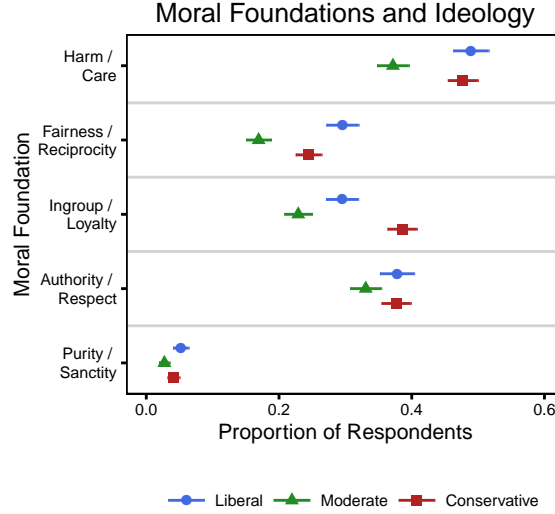
Figure 1: Weighted proportion of Respondents mentioning each of the moral foundations in any of their open-ended responses, along with 95% confidence intervals.

fairness/reciprocity foundation. This pattern is consistent with MFT, as is the tendency for a greater proportion of conservatives to reference the ingroup/loyalty foundation than liberals or moderates. There were some notable contradictions, however. Fewer liberals used fairness/reciprocity words than authority/respect words. Indeed, the proportion of liberal respondents referencing authority/respect is almost identical to the proportion of conservatives mentioning authority/respect.

The fact that the purity/sanctity foundation was almost never mentioned by any of the respondents is surprising, since other studies found that the purity/sanctity foundation plays a very important role when looking at ideological differences (Koleva et al., 2012). This result suggests that subsequent analyses of survey responses might necessitate a revision of the moral foundation dictionary, since the terms contained in the dictionary might not be relevant for political evaluations. Accordingly, some of the words (e.g. in the case of purity/sanctity) are too uncommon in the political context. Due to the very rare mentioning of the purity/sanctity dimension, the subsequent analyses will concentrate on the remaining four moral foundations.[9]

Overall, Figure 1 provides an inconclusive picture on the moral foundations of political reasoning. While most patterns are consistent with our theoretical expectations, the findings

---

[9]Unfortunately, this issue cannot not be properly addressed by relying on tf-idf cosine scores. The weighting can correct for some distortions due to individual ubiquitous terms in the dictionaries, but it cannot compensate for the fact that the purity dictionary as a whole only contains words that are almost never mentioned by respondents.

for the authority/respect dimension contradict MFT. However, part of the inconsistencies might be attributed to the limitations of the dictionaries discussed above (i.e. ubiquitous words and differences in individual response length). In order to alleviate this concern, we now turn to the analyses of the similarity scores as measures of moral reasoning.

I begin by estimating a set of Tobit regressions using ideology (and the control variables discussed above) to predict the individual similarity score for each of the moral foundations (excluding purity/sanctity).[10] Figure 2 compares liberals and conservatives while holding all other variables constant at their respective means. The effects of ideology (liberal - conservative) are decomposed into two parts: the left panel displays the change in the probability to mention a specific foundation, whereas the right panel displays the expected change in the similarity score among respondents who mentioned the foundation (i.e. given that the similarity score is larger 0).
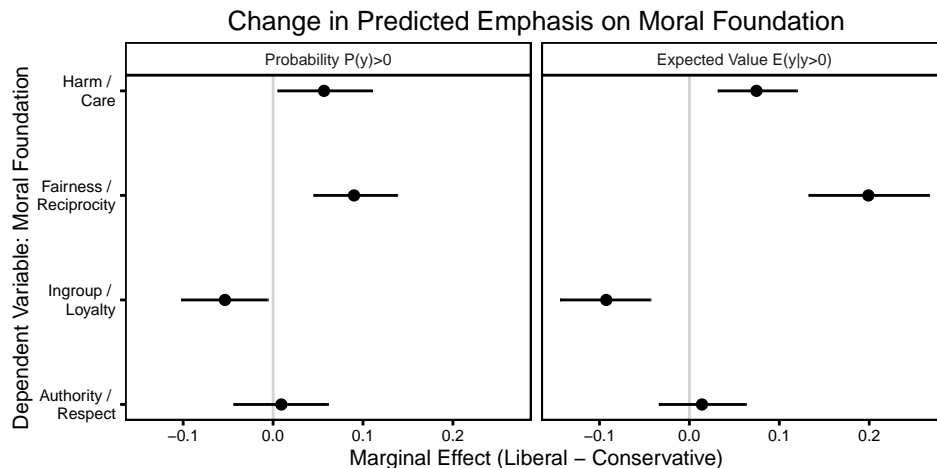

Figure 2: Change between liberals and conservatives in the probability to mention each moral foundation (left panel), and in the similarity score given that the foundation was mentioned (right panel), holding all other control variables at their respective means (along with 95% confidence intervals). Positive values indicate that liberals are more likely to mention the respective foundation (left panel), or emphasize it more than conservatives (right panel), and vice versa. Estimates are based on Tobit models for each foundation's similarity score.

Positive values denote a higher similarity score for the respective moral foundation of a response among respondents who identified as liberals, while negative values indicate a higher score among conservatives. The effects are consistent with the first hypothesis for three out of four moral foundations. Liberals emphasize the foundations of harm/care and fairness/reciprocity significantly more than conservatives when evaluating political parties

---

[10] The full estimates for this and all subsequent models discussed the remainder of the paper are presented in Appendix D.

and candidates. The similarity score for the harm/care foundation is about 0.15 standard deviations higher among liberals than conservatives. The effect is slightly larger for the fairness/reciprocity dimension. Conversely, being conservative increases the similarity score for the foundation of ingroup/loyalty by about 0.1 standard deviations. For the authority/respect dimension, we now see an effect in the expected direction (i.e., conservatives have higher similarity scores). However, the estimated difference between liberals and conservatives on this dimension is small and does not reach conventional levels of significance.

Taken together, the results indicate that there are important differences between liberals and conservatives in their reliance on different moral considerations when evaluating political parties and candidates. However, the patterns are not unequivocally consistent with the predictions of MFT. The fact that some foundations showed insignificant patterns, as well as the failure to replicate the findings across years suggests that the reliance on moral foundations might be more context-specific than previously theorized.

3.3.2 Consequences and Political Relevance of Moral Reasoning

Previous research relying on the MFQ has linked moral foundations to an array of political outcomes, such as turnout (Johnson et al., 2014), candidate preferences (Iyer et al., 2010), and voting behavior (Franks and Scherr, 2015). But do we see the same patterns for explicit moral reasoning as compared to latent moral foundations?

As a first step, we examine the relationship of moral reasoning and political preferences themselves. Figure 3 presents the results of linear regressions of similarity scores for moral foundations predicting the change in the feeling thermometer differential between the Republican and the Democratic party (top part of the figure) and the change in the feeling thermometer differential between both Presidential candidates. Positive values indicate more favorable evaluations for the Democratic candidate or party and negative values indicate more favorable evaluations of the Republican candidate or party. The patterns are largely consistent with the previous results on ideological differences. Individuals who emphasize considerations related to harm/care, and fairness/reciprocity evaluate the Democratic candidates on average between 1 and 4 points higher than the Republican candidates (on a 100 point scale). On the other hand, if individuals emphasized the ingroup/loyalty dimension, they reported stronger preferences for the Republican candidates. These effects are robust after controlling for individual party identification. Thus, in both analyses in Figure 3, we observed sizable and significant effects for the influence of moral reasoning.

This result is noteworthy given that respondents were not explicitly asked about morality. Furthermore, it is not necessary to distinguish between responses for either party or between positive and negative statements in order to recover meaningful patterns. Even though we

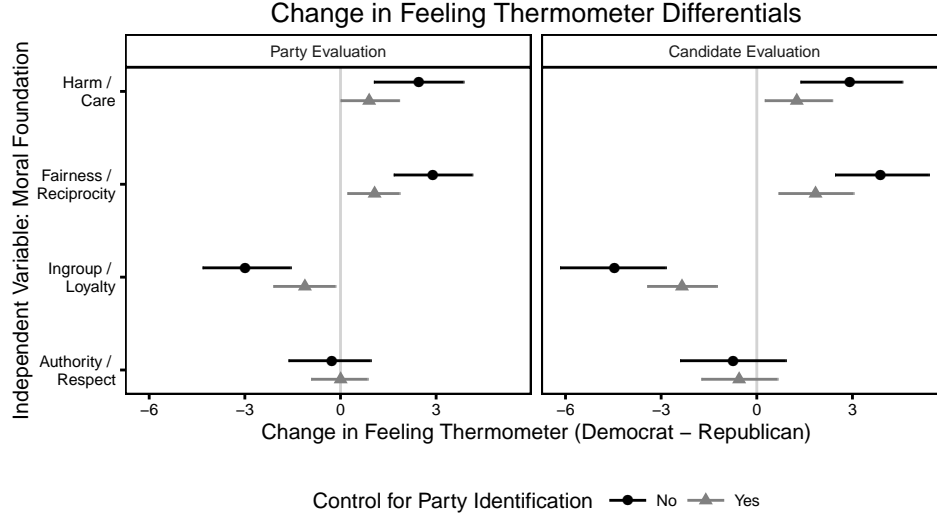## Change in Feeling Thermometer Differentials

Figure 3: Change in predicted feeling thermometer differential when similarity score is increased from its minimum (no overlap between dictionary and response) by one standard deviation, holding all other control variables constant at their respective means (along with 95% confidence intervals). Positive values indicate that respondents who emphasized the respective foundation evaluated the Democratic candidate/party more favorably than the Republican candidate/party, and vice versa. Estimates are based on a single OLS model (using robust standard errors) including similarity scores for each foundation and dotted lines indicate estimates while additionally controlling for party identification.

are simply looking at moral reasoning in the collection of all positive and negative statements about both candidates and parties, we observe consistent and substantial effects on subsequent evaluations. The moral considerations evoked by respondents allow us to make inferences about their political attitudes and behavior irrespective of the specific party or candidate they are evaluating.

Figure 4 presents the changes in expected probabilities of voting for the Democratic (vs. the Republican) presidential candidate in the 2012 election for individuals emphasizing the moral foundations in their open-ended responses. The estimated probabilities are based on logit models including similarity scores for each moral foundation as independent variables as well as several sociodemographic control variables, which were held constant at their mean values when calculating expected values.

Again, the patterns are strikingly similar to the results presented thus far. Individuals who emphasized moral considerations related to the harm/care and fairness/reciprocity foundations are more likely to vote for Barack Obama than for Mitt Romney. Respondents who emphasized the ingroup/loyalty foundation, on the other hand, were less likely to vote for Obama. Again, these results are largely consistent with MFT, although not all effects
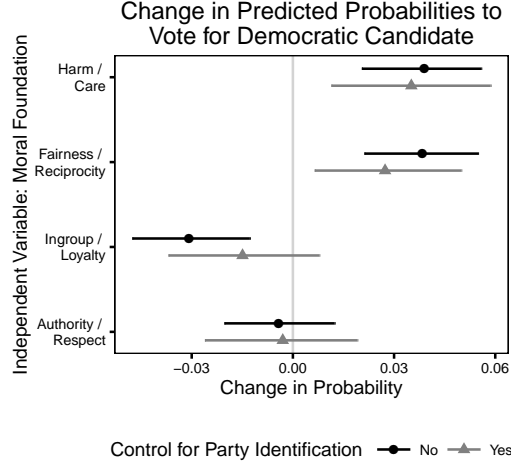
16

Figure 4: Difference in predicted probabilities to vote for Democratic candidate when similarity score is increased from its minimum (no overlap between dictionary and response) by one standard deviation, holding all other control variables constant at their respective means (along with 95% confidence intervals). Positive values indicate that respondents who emphasize the respective foundation are more likely to vote for the Democratic candidate, and vice versa. Estimates are based on a single logit model including similarity scores or each foundation and dotted lines indicate estimates while additionally controlling for party identification.

are statistically significant (especially regarding the authority/respect dimension). Overall, Figure 4 shows that it is possible to predict voting behavior simply by observing the moral dimensions of the respondents' political reasoning, without taking into account which candidate they described, and whether the description is framed positively or negatively. Moreover, these effects largely persist after controlling for individual party identification.

These effects might not seem large, but bear in mind that the independent variable consists of a measure of emphasis of moral considerations in open-ended questions. The fact that we can recover consistent and statistically significant effects on political participation in such a context (even after controlling for strength of party identification and the length of individual responses) is therefore quite meaningful.

Overall then, the analyses indicate that moral reasoning is positively associated with political engagement. That said, the analyses presented here are not sufficient to make a strong causal claim about the direction of this relationship. Yet, the results suggest that moral reasoning (as measured by open-ended survey responses) is powerfully related to different forms of political participation and engagement.

3.3.3 Determinants of Moral Reasoning

Having shown that liberals and conservatives differ with regard to the moral foundations they emphasize when evaluating political actors, I next investigate whether the overall reliance on moral considerations is a product of exposure to political discourse an political engagement. One potential explanation for this relationship could be the fact that individuals differ in the degree to which they moralized their political attitudes, which itself could be a product of exposure to political discourse. Therefore, I regress the sum of similarity scores for each individual on political sophistication, political media exposure, and frequency of political discussions. Figure 5 depicts the respective effects when each independent variable (knowledge, media exposure, discussion) is increased from its empirical minimum value to its empirical maximum value, holding all other variables (including the number of words in each individual response) constant at their means.
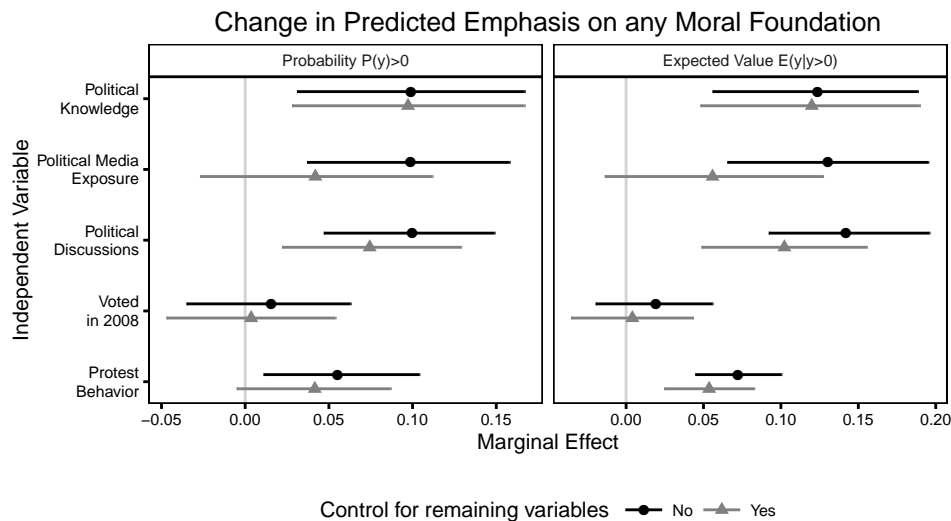

Figure 5: Change in predicted overall reliance on moral foundations depending on political knowledge, media exposure, and frequency of political discussions. The plot shows the predicted difference in the summed similarity scores if each of the independent variables is increased from its minimum to its maximum value holding all other variables constant at their respective means (along with 95% confidence intervals, based on robust standard errors). Positive values indicate higher emphasis of any moral foundation. Estimates are based on OLS and dotted lines indicate estimates while controlling for both remaining variables.

The results show that all three variables have a positive effect on the individual emphasis on moral foundations when evaluating political parties and candidates. Higher political sophistication, higher exposure to political media and news, as well as more frequent political discussions increase the degree to which individuals rely on moral considerations. Thus,

18

citizens *learn* to embed moral reasoning in their political evaluations. While moral intuitions themselves might well be innate, the extent to which individuals make use of these intuitions when thinking about politics and evaluating political actors is context-dependent and subject to individual heterogeneity.

The significant positive effect of frequent political discussions (even after controlling for the two remaining variables, political knowledge and media exposure), is especially interesting in this context. Citizens, who engage in frequent political arguments are more likely to use moral considerations when evaluating candidates and parties. This result suggests that morality serves as a rhetorical tool utilized to convince others of certain political views.

In addition to examining the reference to moral foundations *in general*, we can also consider whether the effects of political discourse increase the differences in the emphasis on specific dimensions between liberals and conservatives. Figure 6 presents the change in the effect of ideology on the individual similarity scores moderated by political knowledge, media exposure, and frequency of political discussions. Thus, each point in the figure displays the interaction effect between knowledge (or media exposure, or discussion) and ideology as a difference-in-difference in predicting similarity holding all other variables constant at their respective means. For example, the positive effect for political knowledge on the similarity score for harm/care (top left part in the figure) indicates, that when political knowledge is increased from its empirical minimum to its maximum, the difference between liberals and conservatives is increased such that the average similarity with the harm/care foundation is about 0.5 standard deviations higher for liberals than for conservatives. Thus, positive effects imply that the gap between liberals and conservatives on the respective dimension is increased in favor of liberal respondents. Negative effects, on the other hand, indicate that the gap between liberals and conservatives is increased with conservatives showing stronger emphasis of the respective moral dimension.

In order to interpret these difference-in-difference effects, consider again the basic findings reported in Figure 2. The estimates showed that on average, liberal responses were more similar to the foundations of harm/care and fairness/reciprocity, whereas conservative responses were more similar to the ingroup/loyalty dimension. Figure 6 shows that for individuals with high political knowledge and media exposure, the difference between liberals and conservatives in terms of the harm/care foundation was more pronounced: liberals are even *more* likely to reference this dimension compared to conservatives. While we do not observe any meaningful moderation effects for the fairness/reciprocity dimension, there is some evidence that the ideological gap in the ingroup/loyalty dimension is higher for respondents who discuss politics more frequently.
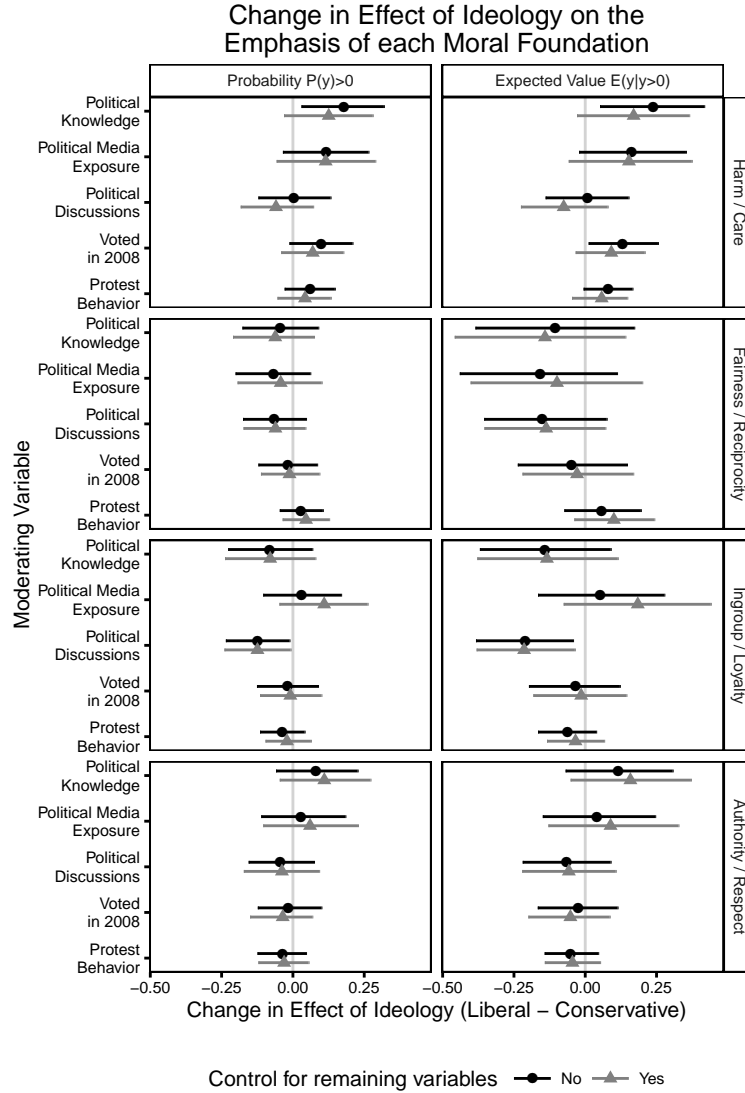
19

Figure 6: Change in effect of ideology on emphasis of each moral foundation moderated by political knowledge, media exposure, and frequency of political discussions (difference-in-difference). The plot shows how the difference in predicted similarity scores for each moral foundation between liberals and conservatives changes if each of the independent variables is increased from its minimum to its maximum value holding all other variables constant at their respective means (along with 95% confidence intervals). Positive values indicate that liberals are more likely to mention a specific moral foundation if they score high on the moderating variable (knowledge, exposure, or discussions), and vice versa. Estimates are based on individual OLS models for each foundation and dotted lines indicate estimates while controlling for both remaining variables.

Overall, political knowledge, media exposure, and political discussions do not only affect general levels of moral reasoning but also moderate the ideological gap between liberals and conservatives. This finding indicates that the effects of political discourse cannot be reduced to an artefact of differences in political literacy and more elaborate argumentation. Some portion of the ideological differences in emphasis on moral foundations can therefore be described as a product of learning in the political environment.

4   Conclusion

The goal of this paper was to investigate whether the ideological differences in the emphasis of moral foundations manifests itself in individual reasoning about political actors. The analyses of open-ended survey responses can provide important insights beyond previous research because it allows us to evaluate whether citizens make references to moral considerations in a political context that does not induce an explicit connection to morality. In contrast to previous accounts of MFT, I argue that the reliance on moral reasoning as such is moderated by political knowledge, media exposure, and political discussions. The emphasis of specific moral foundations, in turn, was expected to influence important political outcomes such as candidate evaluations and voting behavior. Furthermore, general tendencies to take into account moral considerations should affect different measures of political participation.

The empirical evidence discussed in this paper extends and qualifies previous research on moral foundations and ideology. The first hypothesis, which predicted systematic patterns in the emphasis on moral considerations among liberals and conservatives, was supported for three out of four foundations. Liberals are more likely to mention considerations related to harm/care and fairness/reciprocity when discussing their political preferences, whereas conservatives are more likely to emphasize the moral foundation of ingroup/loyalty. Evidence for the authority/respect dimension was tentatively consistent with our expectations, but not statistically significant.

According to the second hypothesis, political knowledge, discussions, as well as media consumption increase the reliance on moral considerations. Thus, the evidence suggests that moral reasoning is part of a broader political learning process. At least in some cases, this learning process appears to imply an increased differentiation between liberals and conservatives in terms of the focus on specific foundations as described by MFT.

The last part of the analyses focused on the political relevance of moral reasoning as conceptualized by open-ended survey responses. The results here revealed consistent relationships between individual moral foundations and political preferences and voting behavior. Furthermore, general moral references consistently increase the likelihood of participating in

political action. As such, moral reasoning measured using open-ended survey responses is a politically meaningful and influential concept.

The contributions of this paper are therefore twofold. This study adds to the existing literature on moral foundations by providing new insights into the mechanisms underlying its relationship with ideology. From a general methodological perspective, the paper emphasizes the potential benefits of incorporating open-ended survey responses in research focusing on the determinants and structure of ideology and political reasoning. The paper shows that one can directly assess moral reasoning in surveys that do not contain the MFQ or related measures, simply by relying on open-ended survey responses.

However, it should be noted that many of the relationships discussed in the paper could not be recovered in a replication using data from the 2008 ANES. While the basic patterns persist, many effects in 2008 fail to reach statistical significance. This issue can partly be explained by the fact that the open-ended responses in 2008 were recorded by some interviewers as indirect summaries rather than direct verbatim responses. As such, the open-ended survey data for 2012 is better suited to identify moral references through the respondent's open-ended responses.

The results presented here provide several directions for future research. The fact that purity/sanctity was almost never mentioned as well as the inconsistent effect of ideology on the authority/respect dimension can be attributed to the fact that the moral foundations dictionary was originally applied to the analyses of sermons. Accordingly, subsequent analyses could revise the dictionary in order to make it more applicable for the analyses of survey responses. An alternative approach to the analysis of open-ended survey responses could be the implementation of structural topic models as described by Roberts et al. (2014): instead of using explicit word lists to identify moral reasoning, it would possible to identify specific topics in open-ended responses that are consistent with the moral foundations described by Haidt and Joseph (2008, see also Lin, Xing, and Hauptmann 2008).

One important issue that remains unresolved is the question of causality. Graham, Haidt, and Nosek (2009) rightfully stated that the causal nature of the relationship between moral foundations and ideology is not yet established. Their study did not settle whether individuals first identify as liberal or conservative and then adapt their respective moral judgments, or whether moral considerations shape and structure subsequent ideological thinking itself. Recent research focusing on elite influences on moral reasoning suggests that elite rhetoric plays an important role in shaping individual moral judgment (see for example Clifford and Jerit, 2013; Clifford, Jerit, Rainey, and Motyl, 2015). However, directly examining the underlying causal mechanism requires additional research where moral reasoning is manipulated experimentally.

It would also be worth investigating whether the patterns regarding moral reasoning change over longer periods of time using this method in order to further establish the context-specific nature of moral reasoning. At this time, however, full-text open-ended survey responses are not available in the ANES prior to 2008.

Overall, the analyses of open-ended survey responses can provide important and valuable insights in the context of moral foundations and the individual underpinnings of political ideology. Utilizing available responses to open-ended survey questions provides a useful and still largely neglected data source to investigate political reasoning.

References

Clifford, Scott. 2014. "Linking Issue Stances and Trait Inferences: A Theory of Moral Exemplification." *Journal of Politics* 76 (3): 698–710.

Clifford, Scott, and Jennifer Jerit. 2013. "How Words Do the Work of Politics: Moral Foundations Theory and the Debate over Stem Cell Research." *The Journal of Politics* 75 (03): 659–671.

Clifford, Scott, Jennifer Jerit, Carlisle Rainey, and Matt Motyl. 2015. "Moral Concerns and Policy Attitudes: Investigating the Influence of Elite Rhetoric." *Political Communication* (forthcoming): 1–38.

Clifford, Scott, Vijeth Iyengar, Roberto Cabeza, and Walter. Sinnott-Armstrong. 2015. "Moral Foundations Vignettes: A Standardized Stimulus Database of Scenarios Based on Moral Foundations Theory." *Behavior Research Methods* forthcoming: 1–37.

Converse, Philip E. 1964. "The Nature of Belief Systems in Mass Publics." In *Ideology and Discontent*, ed. David E. Apter. New York: Free Press.

Day, Martin V, Susan T Fiske, Emily L Downing, and Thomas E Trail. 2014. "Shifting liberal and conservative attitudes using moral foundations theory." *Personality and Social Psychology Bulletin* 40 (12): 1559–1573.

Federico, Christopher M, Christopher R Weber, Damla Ergun, and Corrie Hunt. 2013. "Mapping the connections between politics and morality: The multiple sociopolitical orientations involved in moral intuition." *Political Psychology* 34 (4): 589–610.

Feldman, Stanley, and John Zaller. 1992. "The political culture of ambivalence: Ideological responses to the welfare state." *American Journal of Political Science* 36 (1): 268–307.

Franks, Andrew S, and Kyle C Scherr. 2015. "Using Moral Foundations to Predict Voting Behavior: Regression Models from the 2012 US Presidential Election." *Analyses of Social Issues and Public Policy* 15 (1): 213–231.

Goren, Paul. 2001. "Core principles and policy reasoning in mass publics: A test of two theories." *British Journal of Political Science* 31 (1): 159–177.

Goren, Paul. 2004. "Political sophistication and policy reasoning: A reconsideration." *American Journal of Political Science* 48 (3): 462–478.

Graham, Jesse, Brian A Nosek, Jonathan Haidt, Ravi Iyer, Spassena Koleva, and Peter H Ditto. 2011. "Mapping the moral domain." *Journal of personality and social psychology* 101 (2): 366.

Graham, Jesse, Jonathan Haidt, and Brian A. Nosek. 2009. "Liberals and conservatives rely on different sets of moral foundations." *Journal of personality and social psychology* 96 (5): 1029-1046.

Graham, Jesse, Jonathan Haidt, Sena Koleva, Matt Motyl, Ravi Iyer, S Wojcik, and Peter H Ditto. 2013. "Moral foundations theory: The pragmatic validity of moral pluralism." *Advances in experimental social psychology* 47: 55–130.

Haidt, Jonathan. 2001. "The emotional dog and its rational tail: a social intuitionist approach to moral judgment." *Psychological review* 108 (4): 814.

Haidt, Jonathan. 2012. *The righteous mind: Why good people are divided by politics and religion.* Random House LLC.

Haidt, Jonathan, and Craig Joseph. 2008. "The moral mind: How five sets of innate intuitions guide the development of many culture-specific virtues, and perhaps even modules." In *The innate mind*, ed. P. Carruthers, S. Laurence, and S. Stic. Vol. 3. Oxford Univ Press.

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

Haidt, Jonathan, Jesse Graham, and Craig Joseph. 2009. "Above and below left–right: Ideological narratives and moral foundations." *Psychological Inquiry* 20 (2-3): 110–119.

Iyengar, Shanto, and Sean J Westwood. 2015. "Fear and Loathing Across Party Lines: New Evidence on Group Polarization." *American Journal of Political Science* 59 (3): 690–707.

Iyer, Ravi, Jesse Graham, Spassena Koleva, Peter Ditto, and Jonathan Haidt. 2010. "Beyond identity politics: Moral psychology and the 2008 Democratic primary." *Analyses of Social Issues and Public Policy* 10 (1): 293–306.

Johnson, Kate M, Ravi Iyer, Sean P Wojcik, Stephen Vaisey, Andrew Miles, Veronica Chu, and Jesse Graham. 2014. "Ideology-Specific Patterns of Moral Indifference Predict Intentions Not to Vote." *Analyses of Social Issues and Public Policy* 14 (1): 61–77.

Jost, John T. 2006. "The end of the end of ideology." *American Psychologist* 61 (7): 651.

Koleva, Spassena P, Jesse Graham, Ravi Iyer, Peter H Ditto, and Jonathan Haidt. 2012. "Tracing the threads: How five moral concerns (especially Purity) help explain culture war attitudes." *Journal of Research in Personality* 46 (2): 184–194.

Lin, Wei-Hao, Eric Xing, and Alexander Hauptmann. 2008. "A joint topic and perspective model for ideological discourse." In *Machine Learning and Knowledge Discovery in Databases.* Springer.

Low, Michelle, and Ma Glenda Lopez Wui. 2015. "Moral Foundations and Attitudes Towards the Poor." *Current Psychology* forthcoming: 1–11.

Manning, Christopher D, Prabhakar Raghavan, Hinrich Schütze et al. 2008. *Introduction to information retrieval.* Vol. 1 Cambridge university press Cambridge.

Marietta, Morgan, and David C Barker. 2007. "Values as heuristics: core beliefs and voter sophistication in the 2000 republican nomination contest." *Journal of Elections, Public Opinion and Parties* 17 (1): 49–78.

McDonald, John F, and Robert A Moffitt. 1980. "The uses of Tobit analysis." *The review of economics and statistics*: 318–321.

Mullen, Elizabeth, and Linda J Skitka. 2006. "Exploring the psychological underpinnings of the moral mandate effect: Motivated reasoning, group differentiation, or anger?" *Journal of personality and social psychology* 90 (4): 629.

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, Christopher Lucas, Jetson Leder-Luis, Shana Kushner Gadarian, Bethany Albertson, and David G. Rand. 2014. "Structural Topic Models for Open-Ended Survey Responses." *American Journal of Political Science* 58 (4): 1064–1082.

Ryan, Timothy J. 2014. "Reconsidering moral issues in politics." *The Journal of Politics* 76 (02): 380–397.

Ryan, Timothy J. 2016. "No compromise: Political consequences of moralized attitudes." *American Journal of Political Science.*

Skitka, Linda J. 2010. "The psychology of moral conviction." *Social and Personality Psychology Compass* 4 (4): 267–281.

Skitka, Linda J, Christopher W Bauman, and Edward G Sargis. 2005. "Moral conviction: Another contributor to attitude strength or something more?" *Journal of personality and social psychology* 88 (6): 895.

Smith, Kevin B, John R Alford, John R Hibbing, Nicholas G Martin, and Peter K Hatemi. 2016. "Intuitive Ethics and Political Orientations: Testing Moral Foundations as a Theory of Political Ideology." *American Journal of Political Science.*

Suhler, Christopher L, and Patricia Churchland. 2011. "Can innate, modular 'foundations' explain morality? Challenges for Haidt's moral foundations theory." *Journal of cognitive neuroscience* 23 (9): 2103–2116.

Appendices

Appendix A   Moral Foundations Dictionary

Sources:
Graham, Haidt, and Nosek (2009), as well as http://www.moralfoundations.org/

Note:
Words with (*) indicate that the word stem rather than the exact word was matched in the open-ended survey responses.

Harm:
safe*, peace*, compassion*, empath*, sympath*, care, caring, protect*, shield, shelter, amity, secur*, benefit*, defen*, guard*, preserve, harm*, suffer*, war, wars, warl*, warring, fight*, violen*, hurt*, kill, kills, killer*, killed, killing, endanger*, cruel*, brutal*, abuse*, damag*, ruin*, ravage, detriment*, crush*, attack*, annihilate*, destroy, stomp, abandon*, spurn, impair, exploit, exploits, exploited, exploiting, wound*

Fairness:
fair, fairly, fairness, fair*, fairmind*, fairplay, equal*, justice, justness, justifi*, reciproc*, impartial*, egalitar*, rights, equity, evenness, equivalent, unbias*, tolerant, equable, balance*, homologous, unprejudice*, reasonable, constant, honest*, unfair*, unequal*, bias*, unjust*, injust*, bigot*, discriminat*, disproportion*, inequitable, prejud*, dishonest, unscrupulous, dissociate, preference, favoritism, segregat*, exclusion, exclud*

Ingroup:
together, nation*, homeland*, family, families, familial, group, loyal*, patriot*, communal, commune*, communit*, communis*, comrad*, cadre, collectiv*, joint, unison, unite*, fellow*, guild, solidarity, devot*, member, cliqu*, cohort, ally, insider, foreign*, enem*, betray*, treason*, traitor*, treacher*, disloyal*, individual*, apostasy, apostate, deserted, deserter*, deserting, deceiv*, jilt*, imposter, miscreant, spy, sequester, renegade, terroris*, immigra*

Authority:
obey*, obedien*, duty, law, lawful*, legal*, duti*, honor*, respect, respectful*, respected, respects, order*, father*, mother, motherl*, mothering, mothers, tradition*, hierarch*, authorit*, permit, permission, status*, rank*, leader*, class, bourgeoisie, caste*, position, komplian*, command, supremacy, control, submi*, allegian*, serve, abide, defere*, defer, revere*, venerat*, comply, defian*, rebel*, dissent*, subver*, disrespect*, disobe*, sediti*, agitat*, insubordinat*, illegal*, lawless*, insurgent, mutinous, defy*, dissident, unfaithful, alienate, defector, heretic*, nonconformist, oppose, protest, refuse, denounce, remonstrate, riot*, obstruct

Purity:
piety, pious, purity, pure*, clean*, steril*, sacred*, chast*, holy, holiness, saint*, wholesome*, celiba*, abstention, virgin, virgins, virginity, virginal, austerity, integrity, modesty, abstinen*, abstemiousness, upright, limpid, unadulterated, maiden, virtuous, refined, intemperate, decen*, immaculate, innocent, pristine, humble, disgust*, deprav*, disease*, unclean*, contagio*, indecen*, sin, sinful*, sinner*, sins, sinned, sinning, slut*, whore, dirt*, impiety, impious, profan*, gross, repuls*, sick*, promiscu*, lewd*, adulter*, debauche*, defile*, tramp, prostitut*, unchaste, wanton, profligate, filth*, trashy, obscen*, lax, taint*, stain*, tarnish*, debase*, desecrat*, wicked*, blemish, exploitat*, pervert, wretched*

Appendix B Overview Open-Ended Responses

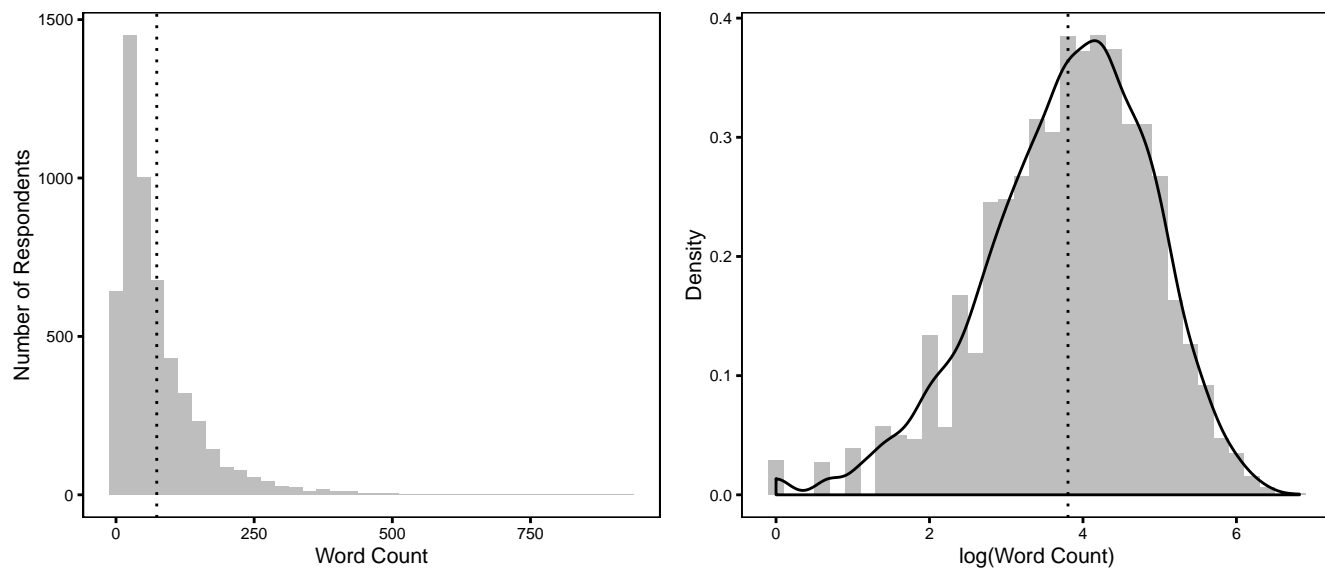|	N	Percent
Spanish Interview	228	3.86
No Responses	417	7.05

Table B.1: Missing open-ended responses

Figure B.1: Histograms displaying the distribution of individual response lengths in number of words for each respective item category. Dotted lines indicate the average response length.

Appendix C Description of Moral Reasoning Measure

This part of the appendix contains a detailed description of the measure of moral reasoning used throughout the paper. It is based on open-ended survey responses as well as the moral dictionary proposed by Graham, Haidt, and Nosek (2009). All techniques described here are adapted from common approaches in quantitative text analysis and information retrieval (see for exampe Manning et al., 2008, for an introduction and from which much of the notation in this part of the appendix is adapted).

As a first step, each individual response as well as the moral foundations dictionaries are converted into a *document-term matrix*. Each row in the document-term matrix represents one document (i.e. the collection of open-ended responses for one individual, or the dictionary for a single moral foundation, respectively). Each column represents a term that is contained in the complete vocabulary of the corpus (i.e. all columns encompass the set of unique words contained in all documents). Each element in the matrix indicates the number of times the term is included in the respective document. As such, the matrix consists of raw frequencies of term t in document d (often denoted by $\text{tf}_{t,d}$).

However, in the context of information retrieval that aims at matching string queries with most relevant matches out of a set of documents, term frequencies are usually weighted by the *inverse document frequency* in order to account for the fact that some terms have more discriminative power than others. The underlying logic is that a word that appears in almost all of the documents should be less decisive in affecting the similarity between the query and the document than a term that only appears in a selection of documents. The inverse document frequency is usually defined as

$$\text{idf}_t = \log \frac{N}{\text{df}_t}, \tag{1}$$

where N is the total number of documents in the corpus, and df_t denotes the number of documents in which term t occurs. The tf-idf score of term t in document d is then computed as

$$\text{tf-idf}_{t,d} = \text{tf}_{t,d} * \text{idf}_t. \tag{2}$$

Raw term frequencies are therefore weighted by the inverse of the term's document frequency, which leads to increased (decreased) values if a term is only included in few (many) documents. We can now represent each moral dictionary and response as a vector of tf-idf scores for each term. For convenience, let $w_{t,d}$ be the tf-idf score for term t in the dictionary and document d (which could be a moral dictionary or an open-ended response). Then,

$$\vec{m}_j = (w_{1,j}, w_{2,j}, \cdots, w_{T,j}) \tag{3}$$
$$\vec{r}_n = (w_{1,n}, w_{2,n}, \cdots, w_{T,n}), \tag{4}$$

where $\vec{m}_j$ denotes the moral dictionary for foundation $j \in \{1, \cdots, J\}$, $\vec{r}_n$ denotes the open-ended responses of individual $n \in \{1, \cdots, N\}$, and T denotes the total number of unique terms in the entire text corpus. This representation of documents and queries (or in our case dictionaries) is usually described as the *vector space model* for representing the corpus.

Each moral dictionary and open-ended response is now represented as a vector of length T. A common measure of relevance, for example of a document ($\vec{r}_n$) for a keyword search query (in our case $\vec{m}_j$), is the cosine similarity:

$$\cos \theta_{j,n} = \frac{\vec{m}(j) \cdot \vec{r}(n)}{|\vec{m}(j)||\vec{r}(n)|}, \tag{5}$$

where $\vec{m}(j) \cdot \vec{r}(n)$ is the dot-product of the dictionary and the response, and $|\vec{m}(j)||\vec{r}(n)|$ is the product of the Euclidean norm of both vectors. $\theta_{j,n}$ can therefore be understood as the angle between the vector space representation of the moral foundation dictionary j and the open-ended response of individual n. The cosine score is 1 if the relative frequency of words used by the respondent overlap completely with the terms in the dictionary (not that this holds irrespective of the length of the response), and 0 if there is no overlap.

Appendix D   Tables of Model Estimates