# Change My View
## Do Moral Appeals Facilitate Compromise?*

Patrick W. Kraft†

– WORK IN PROGRESS –

PLEASE DO NOT CITE OR REDISTRIBUTE WITHOUT PERMISSION

**Abstract**

The American electorate is becoming increasingly polarized. According to research in moral psychology, these growing disagreements between liberals and conservatives can be attributed to fundamental differences in the moral frameworks that shape individual ideology. Indeed, scholars suggest that ideologues would be more likely to reach compromise if both sides spoke the same "moral language." While this implicit assumption has intuitive appeal, it remains largely untested empirically. Drawing on a unique dataset from the online discussion board *Reddit*, this paper examines how moral appeals can affect individual persuasion and the likelihood of compromise.

*Keywords*: Moral Foundations, Conviction, Attitude Change, Persuasion, Compromise

*Previous versions of this manuscript have been presented at the Social Justice Lab (NYU), the Nam Lab (Stony Brook University), and APSA 2018. I thank Jennifer Jerit, John Jost, Yanna Krupnikov, Hannah Nam, Michael Peress, Arthur Spirling, and all seminar participants for helpful comments. The manuscript and code are available on GitHub: https://github.com/pwkraft/cmv.

†Assistant Professor, University of Wisconsin-Milwaukee, kraftp@uwm.edu.

Recent years have witnessed a resurgence in partisan polarization in the United States. Politically engaged citizens hold more diverging policy views, are more ideologically extreme, and exhibit stronger negative affect towards out-partisans than in the past (Hetherington, 2001; Abramowitz and Saunders, 2008; Iyengar, Sood, and Lelkes, 2012; Mason, 2015; Huddy, Mason, and Aarøe, 2015; Iyengar and Westwood, 2015). A growing literature in moral psychology attributes this divide (at least partially) to fundamental differences in moral frameworks that guide liberal and conservative thinking (e.g., Haidt, 2012; Graham et al., 2013). A recent analysis by Garrett and Bankert (2018), for example, finds that individual tendencies to moralize politics exacerbates affective polarization between Democrats and Republicans, which ultimately results in greater social distance and hostility towards out-partisans. More generally, moral conviction as an attribute of attitude strength has been shown to have wide-ranging behavioral consequences (Skitka, Bauman, and Sargis, 2005; Skitka and Morgan, 2014), including diminishing people's willingness to compromise in the realm of politics (Ryan, 2014, 2017).

Do these findings imply that morality in politics is always bound to foster disagreements and hostility between opposing views? Recent research building on Moral Foundations Theory pioneered by Haidt (2007) and colleagues suggests otherwise. According to this view, disagreements about morality are rooted in the underlying intuitions that form people's moral frameworks (Haidt, 2012). For instance, differential emphasis on basic moral dimensions predicts attitudes on culturally divisive issues such as abortion, the death penalty, or same-sex marriage (Koleva et al., 2012). More importantly, however, speaking the same "moral language" can overcome ideological divides. Indeed, political arguments can persuade individuals holding opposing views to the extent that they are emphasizing common moral ground (e.g., Day et al., 2014; Feinberg and Willer, 2015). Moral frames that rely on this logic, for example, were shown to be effective in convincing conservatives to embrace environmental protection policies and sustainable behavior (Kidwell, Farmer, and Hardesty, 2013; Feinberg and Willer, 2013).

However, few studies examined the persuasiveness of congruent moral appeals beyond the context of simple framing scenarios. Instead, they mostly focus on the effect of isolated mes-

sages without giving participants real opportunities to respond or engage in a dialogue. Political discourse is more complex and it is therefore unclear whether previous findings directly translate into more dynamic environments. Accordingly, the suggested potential of moral arguments to help overcoming disagreements—for example in the context of political discussions—is largely assumed as a potential implication and has not been subjected to a direct empirical test. Political discussions are an important source of information for citizens (Huckfeldt et al., 1995) and they have been shown to increase engagement and tolerance of opposing views (Mutz, 2002). Furthermore, Druckman and Nelson (2003) demonstrate that elite framing effects—often viewed as a potential source of polarization—can be mitigated by discussions in heterogeneous groups. Other research shows that such conversations can overcome polarization and partisanship (Klar, 2014). Notwithstanding, most research on deliberation pays little attention to the actual discussion contents (see Barabas, 2004; Karpowitz, Mendelberg, and Shaker, 2012; Mendelberg, Karpowitz, and Oliphant, 2014, for notable exceptions). As a result, we know very little about the role of moral arguments as a potential moderator of discussion effects, which—depending on the perspective in moral psychology—might hurt or harm the potential for compromise.

The present study fills this gap by analyzing the content of more than 10,000 conversations on the active *Reddit* community `/r/ChangeMyView`[1] (CMV). Discussions on CMV—which are anonymous but at the same time successful in maintaining civil discourse—provide an ideal environment to explore correlates of argument persuasiveness across a wide array of topics. For the analyses presented here, I rely on a dataset of matched argument pairs extracted from CMV by Tan et al. (2016), who focused on the role of linguistic features that predict argument strength. My analysis extends these results by examining the effects of moral appeals on attitude change. The findings show that moral arguments can facilitate compromise, but only to the extent that they are congruent with the moral framework of the opposing discussant.

[1] https://www.reddit.com/r/changemyview/

# 1 Theoretical Background

Politics is centered around persuasion and the exchange of opposing arguments. Officeholders, legislators, and activists spend much of their time trying to convince citizens to support one policy over another. As Cobb and Kuklinski (1997) eloquently note, "[p]ersuasion, changing another's beliefs and attitudes, is about influence; and influence is the essence of politics" (88-89). Of course, attempts to persuade are not only limited to elite communications. Citizens discuss political issues with their peers, which turns social networks into a major information source influencing individual attitudes (e.g., Huckfeldt et al., 1995; Ahn, Huckfeldt, and Ryan, 2010; Lazer et al., 2010). The following sections briefly discuss previous approaches to persuasion in politics and connects them to research in moral psychology that helps inform our understanding of the nature of compelling arguments.

## 1.1 Two Routes to Persuasion

One influential framework to conceptualize and explain persuasive communication is the Elaboration - Likelihood Model (ELM) developed by Petty and Cacioppo (1986a,b). The theory distinguishes two separate routes to persuasion, each characterized by their distinctive consequences for a message's effectiveness to change people's attitudes. The first type—the *central route*—is a result of thoughtful processing and a thorough evaluation of the argument's merit. According to this process, people who are sufficiently motivated will incorporate arguments after careful consideration and update their attitudes accordingly. The second type of persuasion, on the other hand, does not require elaborate processing but rather relies on simple cues based on the source of the argument (e.g., group membership, attractiveness, etc.). This route to persuasion is called the *peripheral route* and it can operate without much scrutiny regarding the content of the message (see also Chaiken and Eagly, 1989, for a similar distinction between systematic and heuristic processing). It follows from this distinction that people's motivation and capability to engage in elaborate processing determines whether the persuasiveness of communications is

driven by argument strength itself or rather peripheral cues.

Since contextual factors and individual predispositions affect whether messages are closely scrutinized, different types of arguments may be more or less effective under varying circumstances. For example, Cobb and Kuklinski (1997) analyze the influence of an argument's complexity on it's persuasiveness in two issue areas (NAFTA and health care). Interestingly, they find that while complex arguments were more compelling in the context of international trade, simple arguments proved more effective when discussing the issue of health care. However, the question of why these differences arise is left largely unanswered by Cobb and Kuklinski (1997). One explanation for the inconsistencies is the variation in people's motivation and ability to engage in more thoughtful processing (i.e., their elaboration likelihood). In the absence of such motivation, they are more likely to rely on peripheral cues which renders complex arguments ineffective. A potential motivating stimulus may be the argument's linkage to a person's values. For example, Nelson and Garst (2005) presents experimental evidence showing that people are paying more attention to messages that are consistent with their own value orientation. Participants who received messages that evoked their own values engaged in deeper processing which ultimately made them favor strong arguments and resist weak ones.

Moral appeals may therefore influence the effectiveness of persuasive communications through multiple channels. They may directly improve the merits of the argument itself (central route), they may serve as identity-based cues and heuristics (peripheral route), or they may increase a person's motivation to scrutinize a message in a more elaborate way (see also Petty and Cacioppo, 1986b). As will be further described below, the present analysis focuses on the influence of moral appeals on argument strength in the context of elaborate processing and the central route to persuasion.

## 1.2 Morality and the Potential for Compromise

There are two broad strands of literature in moral psychology that ultimately lead to diverging predictions regarding the effects of moral appeals on argument persuasiveness. Research on *Moral*

*Conviction* conceptualizes moralization as a unique feature of attitude strength (Skitka, Bauman, and Sargis, 2005). According to this view, moral convictions are perceived as "absolutes, or universal standards of truth that others should also share" (Skitka, 2010, 269). As such, moral convictions are viewed by individuals as applying to everyone (universality), they do not require an immediate underlying rationale but are rather seen as facts about the world (objectivity), they can be independent of authority and group norms (autonomy), they elicit strong emotional reactions, and they have an inherent motivational quality (motivation/justification) (Skitka, 2010).

Building on this work, Ryan (2014) argues that moral convictions are not restricted to issues that are traditionally perceived as "moral," such as abortion or same-sex marriage, but can also include other issues such as economic policies. The degree of moral conviction may therefore vary between individuals as well as across issues. Ryan (2014) further shows that the propensity to moralize—i.e. the tendency to view an issue as a question of "right and wrong"—is related to political participation, extreme political attitudes, arousal of negative emotions, and hostility. In a subsequent study, Ryan (2017) suggests that moralization reorients behavior from maximizing gains to the general adherence to rules. Across multiple studies, the author shows that this tendency translates into stronger opposition to compromise about political issues and decreased support for compromising politicians. These patterns should also translate into attitudes towards—and interactions with—others who hold opposing views. Indeed, moral conviction has been shown to be related to stronger preferences for social distance from (and hostility towards) attitudinally dissimilar others and lower cooperativeness in groups holding heterogeneous views (Skitka, Bauman, and Sargis, 2005). This theoretical perspective therefore ultimately suggests that arguments that emphasize an issue in terms of deeply held moral mandates should entrench people to maintain their prior attitudes and therefore reduce the argument's persuasiveness.

However, not everyone agrees with this general prediction. In fact, *Moral Foundations Theory* (MFT) offers a more differentiated view regarding the role of moral appeals in facilitating compromise. The theory proposes a taxonomy of basic moral intuitions that is closely related to ideological thinking. According MFT, liberals focus on *individualizing* moral foundations, which

include care/harm and fairness/cheating. Conservatives, on the other hand, also emphasize the remaining *binding* foundations of loyalty/betrayal, authority/subversion, and sanctity/degradation (Haidt and Graham, 2007; Graham, Haidt, and Nosek, 2009). Differential emphasis on these moral dimensions is systematically related to attitudes towards a wide variety of divisive political issues (e.g. Koleva et al., 2012; Kertzer et al., 2014; Low and Wui, 2015), personality traits like individual social dominance orientation and right-wing authoritarianism (Federico et al., 2013), as well as voting behavior (Franks and Scherr, 2015). Overall, this body of research suggests that liberals and conservatives endorse different moral foundations and that these differences are closely related to political attitudes, evaluations, and behavior.

An implicit assumption made in this literature is that liberals and conservatives would be more likely to come to agreements *if only they focused on the same moral foundations*. For example Haidt (2012, 365) concludes in his book *The Righteous Mind: Why Good People Are Divided by Politics and Religion*: "Once people join a political team, they get ensnared in its moral matrix. They see confirmation of their grand narrative everywhere, and it's difficult—perhaps impossible—to convince them that they are wrong *if you argue with them from outside of their matrix*" (emphasis added). In an different article, Graham, Haidt, and Nosek (2009, 1040) contend that their findings "help explain *why liberals and conservatives disagree on so many moral issues* and often find it hard to understand how an ethical person could hold the beliefs of the other side: Liberals and conservatives *base their moral values, judgments, and arguments on different configurations* of the five foundations."

Several framing studies examining the effects of moral arguments that are congruent with ideological predispositions support this view. For example, binding appeals have been shown to increase recycling behavior among conservatives, whereas individualizing arguments were effective among liberals (Kidwell, Farmer, and Hardesty, 2013). Similarly, Feinberg and Willer (2013) find that pro-environmental frames emphasizing concerns related to the purity dimension reduce attitudinal gaps of conservatives vis-à-vis liberals. Further studies suggest that morally congruent appeals are effective in shifting attitudes of ideologues on various other issues as well (e.g., Day

et al., 2014; Feinberg and Willer, 2015).

Both theories of morality therefore lead to diverging expectations regarding the effect of moral appeals on the potential for compromise: While the moral conviction literature suggests that *any* type of moral appeal should make it harder to overcome disagreements, MFT contends that agreement can be facilitated if two discussants focus on the same underlying moral dimensions. The question whether emphasizing the same foundations can facilitate compromise has important implications—especially in our current political environment. Somewhat surprisingly, however, this claim has not been subjected to a direct empirical test in the context of political discussions.

## 1.3 Hypotheses

The structure and dynamics of political discussions can be prohibitively complex, making it difficult to derive clear expectations regarding the persuasiveness of individual arguments and their role in achieving compromise. In order to gain some analytical leverage, consider the following simplified scenario of a conversation between two discussants, $\mathcal{A}$ and $\mathcal{B}$, who disagree on some issue $x$. Suppose further that only $\mathcal{A}$'s opinion is malleable and may change as an outcome of the discussion. $\mathcal{B}$'s own position is firm and she is solely trying to challenge $\mathcal{A}$'s view. The conversation begins with $\mathcal{A}$ making an opening statement describing and defending her opinion—potentially relying on moral justifications. $\mathcal{B}$ then engages in the discussion and may try to persuade $\mathcal{A}$ using either moralized or non-moralized arguments. Of course, $\mathcal{A}$ and $\mathcal{B}$ can continue to respond to each others' statements until either $\mathcal{A}$ changes her opinion or the conversation ends without attitude change. Both theoretical perspectives described in the previous section suggest contrasting hypotheses regarding the persuasiveness of $\mathcal{B}$'s appeals:

*H1 (Moral Conviction)*: Arguments that involve moral appeals will be *less* persuasive than arguments that do not involve moral appeals.

*H2 (Moral Foundations)*: Arguments that involve moral appeals will be *more* persuasive than arguments that do not involve moral appeals, but only if they are congruent with the opening statement's moral framework.

To reiterate, in this unidirectional model of a discussion, only $\mathcal{A}$ stands to maintain or change her prior view, whereas $\mathcal{B}$ attempts to persuade her discussant. Compromise is achieved in this scenario if $\mathcal{B}$ is able to persuade $\mathcal{A}$ to change her attitude. One of the main advantages of this structure is that it enables a clear analytical distinction between statements that are intended as justifications to defend and bolster one's own view (i.e., $\mathcal{A}$'s arguments) and challenges that are targeted to alter a discussant's opinion (i.e., $\mathcal{B}$'s arguments), which is not feasible in a free flowing discussion where—at least potentially—all views are malleable. The following section illustrates how conversations on the Reddit community `/r/ChangeMyView` resemble this stylized conceptualization of a discussion and therefore provide an ideal environment to test both competing hypotheses.

## 2 The Subreddit "ChangeMyView"

Reddit is an online discussion board organized into thematic forums called *subreddits*. Users can join these communities based on their interests and each subreddit has its own norms and etiquette that are enforced by voluntary moderators. `/r/ChangeMyView` (CMV) is a subreddit where participants can initiate a discussion by posting an opening statement establishing a personally held view on a particular issue (e.g., "CMV: I believe that the gay marriage discussion isn't as important as the media portrays it to be."), followed by a brief explanation of their underlying rationale. Other users are then invited to challenge the original poster's (OP) opinion by providing counterarguments. OPs respond to the challenges and—crucially—identify individual posts that changed their mind by awarding a "Delta" ($\Delta$). The community is dedicated to civil discourse—even for divisive issues—and encourages OPs to be open to changing their views and to award $\Delta$s genuinely (see also Jhaver, Vora, and Bruckman, 2017).[2] To date, the subreddit has more than 500,000 subscribed users.

As an illustrative example, consider the following discussion on marriage equality that was

[2]The current set of rules for original posts as well as responses can be accessed at https://www.reddit.com/r/changemyview/wiki/rules. Additionally, an overview of the current rules is included in Appendix A.

posted in 2014. The thread begins with the following opening statement (the posts were slightly edited for readability):

> *CMV: I believe that the gay marriage discussion isn't as important as the media portrays it to be.*
>
> The real problem is the concept of marriage itself. In my view, LGBT couples are already married, regardless of the legislation that is imposed on them. Marriage isn't a set of civil rights that confirms your connection to your partner, it's the choice you make to be in private, daily, lifelong commitment to another being.
>
> Tradition dictates that in order to be 'properly' married you have to exchange vows, get a ring, and have a massive celebration (the set of traditions change based upon the culture.) but marriage isn't that, it is simple commitment to another person. The main issue that gay marriage has is that not all couples are given the same civil liberties, but this does not mean that their marriages are void. Marriage isn't decided by bystanders, it's decided by the people who live inside the union. It is for this very reason that a gay couple getting married doesn't affect your own marriage.
>
> I've held this opinion for a while but have never had the opportunity to see if it stood up to criticism. CMV.

Here, the OP argues that marriage equality should be less of a controversy since the defining feature of marriage is the commitment in a relationship rather than its legal status. Several users argued against this view from various perspectives. Below is a sample response that ultimately lead the OP to award a $\Delta$ to indicate that it changed his or her view:

> That would be true if it was just some odd tradition. But it isn't just the ceremony, but also a tax.
>
> Right now there is a gay tax. Gay couples have to pay higher taxes than straight couples because the government gives a tax break for married couples. The reason for this is that married couples tend to be more efficient and better for the government. The government wants to encourage marriage, so as with all things they encourage they subsidize it.
>
> Gay people provide the exact same benefits to marriage, if not more! Adoption being the largest one.
>
> This tax comes through in multiple ways. The yearly tax and through inheritance. The government doesn't tax inheritance as much for marriage, but if they are simply partners then they get taxed when their "partner" dies.
>
> The state also doesn't allow for gay couples see their loved ones in hospitals or prison because they aren't married.
>
> If this was just in the church I wouldn't care. But this is much more than that.

Note that in principle, the OP could have reacted to this root response by providing additional justifications and the discussion between both users could have continued for a few posts. In this case, the OP directly provided a $\Delta$. However, other discussants were less successful in persuading the OP. In contrast to the previous example, the following response did not receive a $\Delta$:

> If gay marriage is not allowed in a state
>
> 1. Their marriages technically *are* null and void, as the state does not recognize them.
> 2. Marriage is not actually decided by the people in the union, since there are legal requirements as well as legal benefits. Which brings me to my next point.
> 3. There are several legal benefits (as well as tax benefits) to being married. States which do not allow gay marriage do not give these legal benefits to gay couples.
>
> You might believe you are married to someone, but the term "marriage" is a political one indeed since it has legal ramifications.

While both responses emphasize the importance of legal considerations in justifying the need for marriage equality, only one of the contributions persuaded the OP sufficiently such that she awarded a $\Delta$.

This online format provides an ideal opportunity to explore the correlates of argument persuasiveness consistent with the stylized structure outlined in the previous section. Discussions begin with a short explanation of a person's opinion on a given topic. Multiple users attempt to counterargue the OP's point of view from various perspectives in a civil dialogue. Most importantly, OPs explicitly identify and label arguments they deemed persuasive enough to change their views. The nature of the conversations on CMV as well as the anonymity of individual users turns the focus on the content and merits of arguments (i.e., the central route to persuasion) rather than source cues and identity-related factors. In contrast to past framing studies which usually implement single messages, users on CMV evaluate a multitude of available arguments, which allows for a unique counterfactual design to study persuasive messages that can be directly linked to the OP's initial justifications. Lastly, examining discussions on CMV allows us to explore a wide array of issues.

Recent research in machine learning and computational linguistics has started to use CMV to study online discussions (Wei, Liu, and Li, 2016; Hidey et al., 2017). The following analyses leverage a set of matched argument pairs extracted from CMV by Tan et al. (2016), who explore interaction dynamics on CMV by analyzing linguistic features (such as, for example, the use of personal pronouns) that predict persuasiveness as well as the malleability of original posts. Their dataset includes more than 10,000 discussions that were posted on the subreddit between January 2013 and May 2015. It is important to note that the analysis published by Tan et al. (2016) focuses less on the content of discussions (i.e., *what is being said*) but rather examined discussion dynamics and linguistic characterics (i.e., *how it is expressed*) to predict persuasiveness. The following analyses explicitly turn to the effects of moral content on discussion outcomes.

# 3 Opinion Change in Online Discussions

Consider again the simplified model of a discussion between person $\mathcal{A}$ and $\mathcal{B}$, where $\mathcal{A}$ stands to defend her view against the challenges put forward by $\mathcal{B}$. While the hypotheses specified above are focused on the persuasiveness of $\mathcal{B}$'s arguments (i.e., discussion posts that respond to the OP in the context of CMV), it is helpful to focus first on the opening statements initiating each discussion and examine the extent to which OPs are willing to award $\Delta$s in the first place.

To provide an initial overview of the range of topics that are covered in the set of 10,000 initial statements included in the data, I extract 20 clusters of co-occurring terms via Latent Dirichlet Allocation (Blei, Ng, and Jordan, 2003). The topic model is solely based on contents of the original posts starting each discussion thread (disregarding subsequent comments). Statements were pre-processed by removing numbers, punctuation, symbols, hyphens, URLs, as well as stopwords. All remaining terms were stemmed and only included if they appeared in at least 10 different posts. Figure 1 presents the average topic proportions across opening statements based on the model. For each topic, the plot additionally displays the ten most likely word stems as well as a descriptive label on the y-axis.

Conversations on CMV range across a variety of topics such as economic issues, gender/sexuality, or domestic and international politics. Of course, it could be argued that some of these topics—for example those related to religion—more easily lend themselves to concerns about morality. Notwithstanding, recent work in moral psychology by Ryan (2014) and others routinely emphasizes that in principle, any issue bears the potential to be moralized by individuals. However, in an effort to preclude any concerns about potential confounding effects related to topic selection, the main analyses reported below focuses on comparing arguments *within* a given discussion thread.

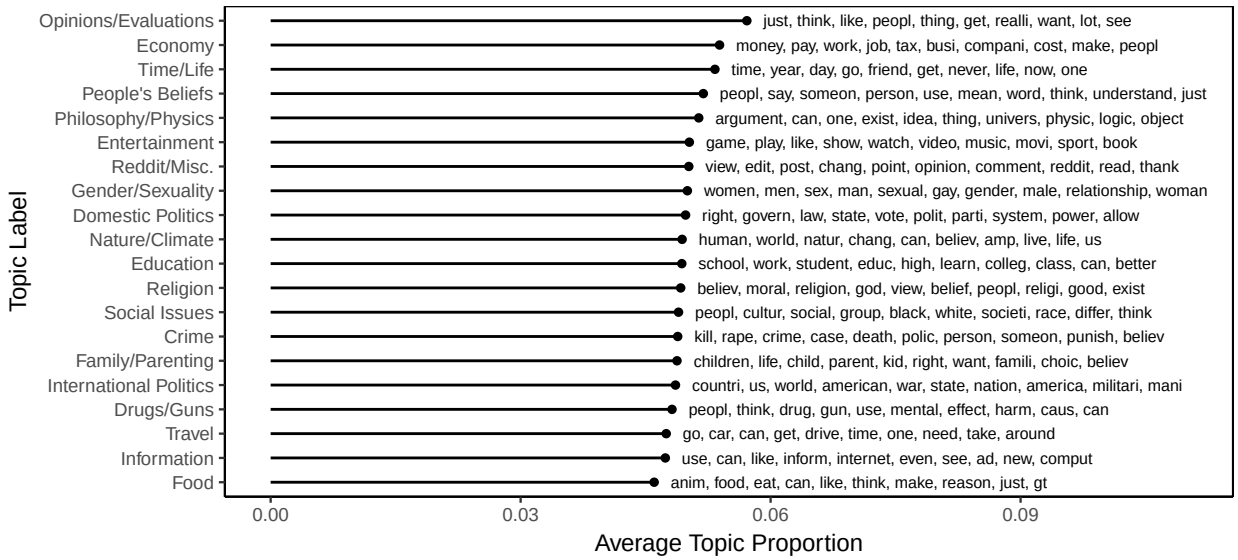


Figure 1: Average topic proportions in opening statements on `/r/ChangeMyView/` based on a basic LDA model with 20 topics. The plot additionally displays the ten most likely terms associated with each respective topic.

The Internet is not necessarily known as a place where people are willing to change their mind about *any* issue. Yet, CMV maintains an open atmosphere that encourages users to acknowledge arguments that change their perspective. The rules of the subreddit state that users should "Award a delta if you've acknowledged a change in your view. [...] Please note that a delta is not a sign of 'defeat', it is just a token of appreciation towards a user who helped tweak or reshape your opinion. A delta also doesn't mean the discussion has ended."[3] Of course, this does not imply that every OP awards a $\Delta$ throughout a conversation. Figure 2 displays the number of

[3] https://www.reddit.com/r/changemyview/wiki/rules, last accessed April 22, 2018

discussion threads included in the dataset where OPs indicated that one of the responses changed their mind.

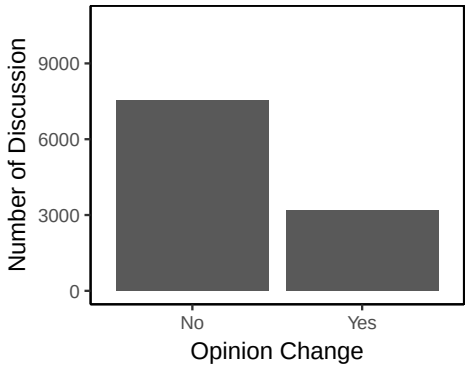


Figure 2: Number of discussions on `/r/ChangeMyView/` that resulted in opinion change (at least one $\Delta$ awarded by OP) versus not.

In about two thirds of discussions on CMV between 2013 and 2015, OPs did not award a $\Delta$ for any of the counterarguments that were put forward, which leaves about 3,000 individual threads where OPs indicated that at least one of the responses changed their views. Interestingly, additional analyses included in the supplementary material show that there are only minimal differences in topic proportions between discussion threads that resulted in persuasion versus threads that did not (see Appendix D).

In their original study, Tan et al. (2016) mainly investigated linguistic patterns (e.g., use of personal pronouns) and differences in style (formatting etc.) that predicted resistance to persuasion among OPs. They conclude for instance that "comparative adjectives and adverbs are a sign of malleability, while superlative adjectives suggest stubbornness." The goal of the present analysis, in turn, is to go beyond linguistic patterns that are unrelated to content and explore the role of moral appeals in facilitating or inhibiting compromise. In order to capture moralized arguments, I rely on the MFT dictionary proposed by Graham, Haidt, and Nosek (2009), which contains lists of word stems that signal each of the five moral foundations (care, fairness, loyalty, authority, sanctity) as well as a category of general moral terms.[4]

[4]See Appendix B for the complete dictionary.

Figure 3 displays the percentage of dictionary terms for each foundation in the opening statements initiating a discussion on CMV (in proportion to the total number of words in each post). The plot compares the reliance on moral terms between OPs that subsequently changed their view versus OPs that did not. As an initial observation, it is interesting to note that the distribution of dictionary terms across foundations is strikingly similar to the proportions of moral terms in open-ended responses to the likes-dislikes questions included in the American National Election study (c.f., Kraft, 2018): The most prevalent dimensions are *care* and *authority*, while occurrences of *sanctity* are largely negligible. Observing these similarities is noteworthy since they are suggestive of a common mechanism driving the emphasis on moral considerations when justifying preferences in a public opinion survey as well as in online discussions.

More important for the purposes of this paper is the fact that the percentage of dictionary terms across foundations appears smaller among opening statements that resulted in opinion change than among those that did not. More specifically, OPs who did not award any $\Delta$s in the subsequent discussion put a significantly stronger emphasis on moral considerations related to loyalty and authority ($p < .01$ in both cases after accounting for multiple comparisons using Bonferroni correction). Similar results can be obtained after aggregating all moral dictionary terms in a single category: OPs who were not persuadable on CMV use more words related to morality overall than OPs who indicate that the discussion changed their view ($p < .001$).

At first look, the findings appear consistent with the moral conviction literature, which posits that people who hold moralized attitudes are less willing to compromise and deviate from their prior beliefs (e.g., Skitka, Bauman, and Sargis, 2005; Ryan, 2014, 2017). Yet, there are important issues that make it difficult to draw strong conclusions based on these initial results. First of all, there may be unobserved confounding factors that are related to both the OPs willingness to award $\Delta$s as well as the chosen discussion topic (which could be more or less aligned with moral considerations). The content of opening statements may also induce selection bias in user responses which can impact the nature of their comments and ultimately the productivity of discussions. Furthermore, there is no way of contrasting the potentially diverging impact of

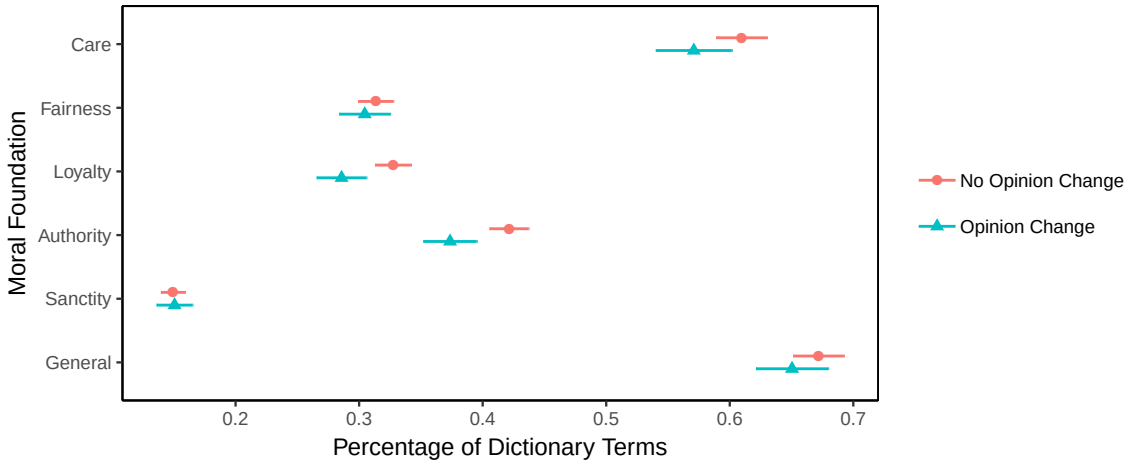


Figure 3: Moral foundations and persuadability: Average percentage of dictionary terms relative to the total number of words in each opening statement beginning a discussion, comparing discussions where the OP subsequently awarded a $\Delta$ (opinion change) or not (including 95% confidence intervals).

morally congruent arguments by exclusively examining the malleability of initial opinions. The following analyses address these problems by comparing the relative persuasiveness of arguments *within a given discussion thread*.

# 4 What Makes an Argument Persuasive?

The previous section demonstrated that the OPs' reliance on moral language in opening statements is inversely related to their willingness to change their view in the subsequent discussion. Now I focus directly on the persuasiveness of comments that are made in response to a given opening statement on CMV. In the context of the simplified discussion framework outlined above, I examine the arguments brought forward by $\mathcal{B}$, who is challenging $\mathcal{A}$'s view. This allows me to directly compare the moral conviction hypothesis with the moral foundations hypothesis, which have diverging predictions regarding the effectiveness of moralized appeals in discussions. Note that the arguments presented by $\mathcal{B}$ do not only include her initial post (i.e., the root response), but also any subsequent posts that are mentioned in the evolving discussion between $\mathcal{A}$ and $\mathcal{B}$ (i.e., the full response path).

In the original analysis by Tan et al. (2016), the authors implement a simple method to select pairs of arguments that respond to the same original post, with only one of the selected responses being successful in changing the OPs view. While differing in persuasiveness, arguments are matched in such a way that they are as similar as possible in terms of their word choice. More specifically, Tan et al. (2016) select argument pairs by maximizing their Jaccard similarity:

$$\text{Jaccard}(B_{\Delta}, B_{\neg\Delta}) = \frac{|B_{\Delta} \cap B_{\neg\Delta}|}{|B_{\Delta} \cup B_{\neg\Delta}|}, \tag{1}$$

where $B_{\Delta}$ and $B_{\neg\Delta}$ are the sets of words in two response paths associated with the same opening statement (one receiving a $\Delta$, the other not). In other words, they match each successful counterargument to an unsuccessful response that shares the largest proportion of common words (disregarding stopwords). As Tan et al. (2016, 617) describe: "This leads to a balanced binary prediction task: which of the two lexically similar rooted path-units is the successful one?" [5]

The analyses reported below rely on this approach to select matched argument pairs for comparison. To reiterate, I focus on discussions in which OPs awarded at least one $\Delta$. A response that received a $\Delta$ is then matched to another argument within the same discussion that was not successful in changing the OP's view but is as similar as possible in terms of its vocabulary. Note that in principle, this strategy should make it more difficult to find differences in the MFT dictionaries as argument pairs are matched based on their lexical similarity. One might worry, however, that the necessary initial selection on discussions where OPs ultimately awarded at least one $\Delta$ might disproportionately discard cases where the initial statement emphasized moral considerations. Luckily, that is not the case. Figure 4 shows that almost all of the opening statements in the matched pair selection mention at least one of the moral dictionary terms. Furthermore, the proportion of moral dictionary terms among this set of opening statement shows the same pattern as in Figure 3 (results included in Appendix C).

An important unresolved issue using this approach is that the matching procedure only focuses

[5] As additional selection criteria and to avoid trivial posts, arguments are removed if they are shorter than 50 words, only include clarifying questions, or if the opening statement received fewer than 10 responses overall and fewer than 3 unsuccessful challenges (see Tan et al., 2016, 617 for details).

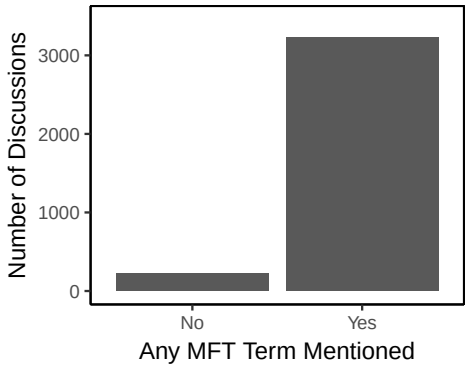


Figure 4: Number of opening statements in the paired argument data that included *any* term mentioned in the MFT dictionary.

on the set of unique words that are used in each response path and does not take into account their relative length. This can be especially problematic since persuasive discussions tend to be longer and involve at least a few back-and-forth exchanges between the OP and the challenger (c.f., Tan et al., 2016, 616). Figure 5 displays the distribution of the differences in word counts between successful and unsuccessful argument pairs. Clearly, longer responses are more likely to be awarded a $\Delta$, which might jeopardize potential inferences about the relative reliance on moral dictionary terms.

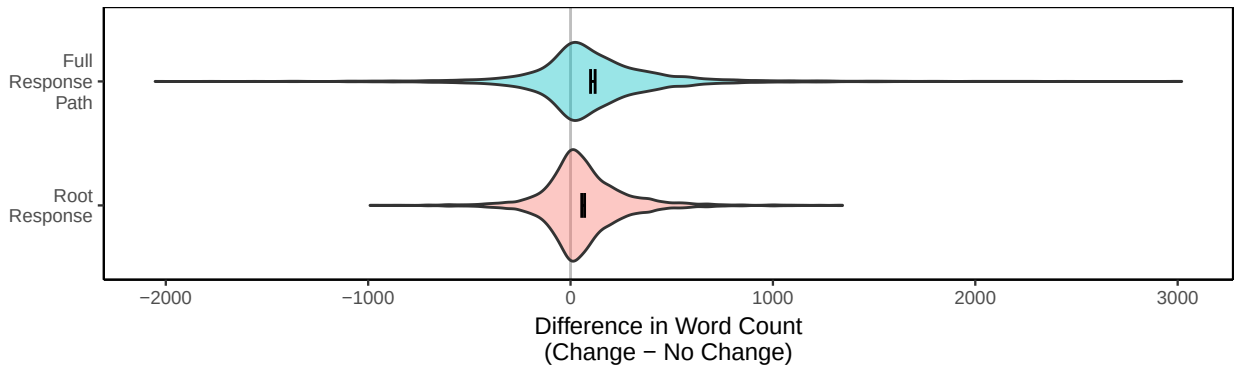


Figure 5: Difference in response lengths between successful and unsuccessful counterarguments. The narrow black bars display the 95% confidence interval of mean differences.

As a first step to alleviate this concern, it is worth noting that I only examine the *percentages* of MFT dictionary occurrences in each discussion rather than a raw count, which implies that the

prevalence of moral considerations is standardized by the overall length of each post. Notwithstanding, I take additional precautions proposed by Tan et al. (2016) to check the robustness of the results. First, I not only examine differences when looking at the entire response path of a discussion between two users (i.e., all posts that were part of the dialogue with the OP), but restrict the analysis to focus on the challenger's root response to the opening statement only. As can be seen in Figure 5, the differences in word counts between argument pairs are significantly smaller. Recovering the same patterns in the root response as in the full response path indicates that the initial arguments that triggered an exchange with the OP are by themselves predictive of the outcome of the discussion. To be fair, there are still substantial differences in the length of successful versus unsuccessful root responses. As a second robustness check, I additionally truncate the longer root response of a given pair as follows: I count the total number of words in each post and simply cut off the end of the longer response such that both word counts in a given pair are exactly equal. While this approach eliminates any concerns about argument length as a confounding factor, it comes at the price of losing a lot of information by ignoring potentially valuable content. Using this framework, I now turn to the analysis of the persuasiveness of moral arguments made in discussions on CMV.

## 4.1 Moral Appeals are Futile...

Recall that the moral conviction hypothesis posits that moralized arguments will be less persuasive than arguments that do not involve moral appeals. In order to test this proposition, I examine the argument pairs matched within discussions and compare MFT dictionary proportions between contributions that were successful in receiving a $\Delta$ and those that were unsuccessful. Figure 6 presents the results of logistic regressions predicting an argument's success in triggering opinion change as a function of moral language use.

The figure displays the marginal effect on the probability of opinion change when increasing MFT dictionary proportions for each foundation from zero to their respective empirical maximum. Positive values indicate that arguments with larger proportions of dictionary terms belonging to

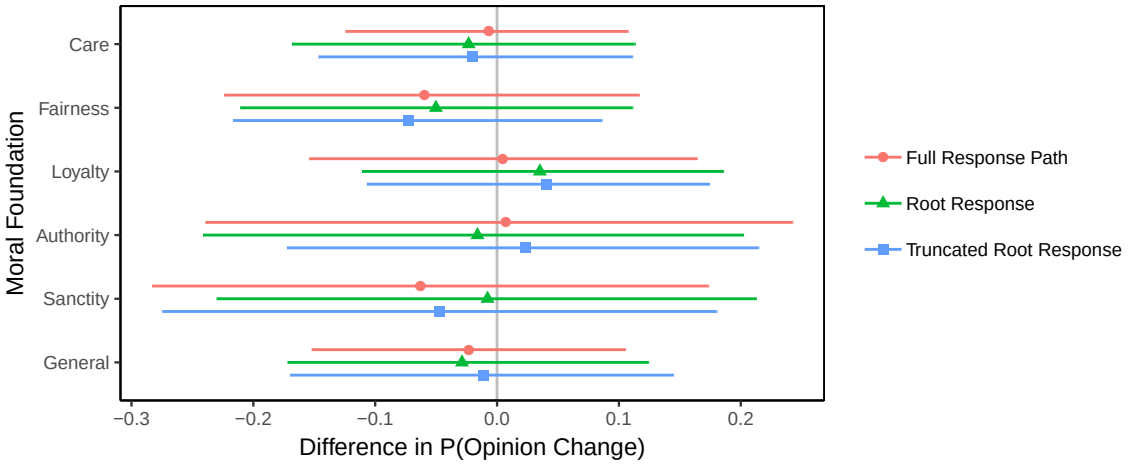


Figure 6: Moral foundations and persuasiveness: Change in predicted probability of opinion change ($\Delta$ awarded) when MFT dictionary term proportions are increased from their minimum (no moral terms mentioned) to their empirical maximum, holding the remaining foundations constant at their mean (including 95% confidence intervals). Estimates are based on logit models with standard errors clustered by discussion thread. Full model results are displayed in the appendix, Table E.1.

a given foundation have a higher probability of receiving a $\Delta$. Again, according to the literature on moral conviction, we would expect the opposite, namely that arguments focusing on moral considerations should be less persuasive. As discussed in the previous section, the analyses are implemented for the full response path as well as focusing only on (truncated) root responses.

The results show that evoking moral considerations in counterarguments does not affect the likelihood of changing the OPs' view on a given issue. This finding furthermore holds after combining all dictionary term proportions in an aggregate measure of moralization across foundations ($p > .45$). Moralized arguments as such are therefore no less persuasive and do not reduce compromise, a result that is not consistent with the moral conviction literature.

## 4.2 ...Unless We're Speaking the Same Moral Language

In contrast to the moral conviction hypothesis, moral foundations theory suggests that we cannot fully understand the effect of moral appeals without taking into account the discussion partner's moral framework. What is decisive from this perspective is the congruence in moral arguments

between both discussants. I measure the moral congruence between an OP's opening statement and each counterargument by computing the cosine similarity in their respective MFT dictionary scores. In general, using cosine similarities based on vectors of word counts is a standard approach in text analysis to quantify the similarity of documents independent of their length (e.g., Manning et al., 2008). More formally, moral congruence can therefore be written as:

$$\text{MFT Congruence} = \frac{\vec{a} \cdot \vec{b}}{||\vec{a}||\,||\vec{b}||}, \tag{2}$$

where $\vec{a}$ is the vector of dictionary counts in the OP's opening statement and $\vec{b}$ is the respective vector for a response. The measure ranges from 0 (no moral overlap) to 1 (equal emphasis on the same moral foundations). Moral congruence is also set to zero if either one of the statements does not contain a single term included in the dictionary.

To reiterate, the moral foundations hypothesis posits that arguments involving moral appeals will be more persuasive than arguments that do not involve moral appeals, but only if they are congruent with the opening statement's moral framework. In contrast, the moral conviction literature would predict a negative effect of moral congruence, since it implies that both discussants, who hold opposing views on an issue, use moralized arguments that ultimately reduce the potential for compromise. Figure 7 displays the effect of moral congruence on argument success. Similar to the previous analysis presented in Figure 6, it shows the predicted change in the probability of persuasion based on a logistic regression model, but now examining the effect of an increase in MFT congruence from its minimum to its maximum. Positive values indicate that posts were more likely to be awarded a $\Delta$ by the OP if they used language that is morally congruent with the OP's opening statement. On average, emphasizing the same moral foundations as the opening statement (as compared to no overlap in moral language at all) increases the probability of opinion change by about 7 percentage points. These results are consistent with MFT as moral congruence is associated with a higher probability of opinion change.

The positive relationship between moral congruence and persuasiveness remains significant

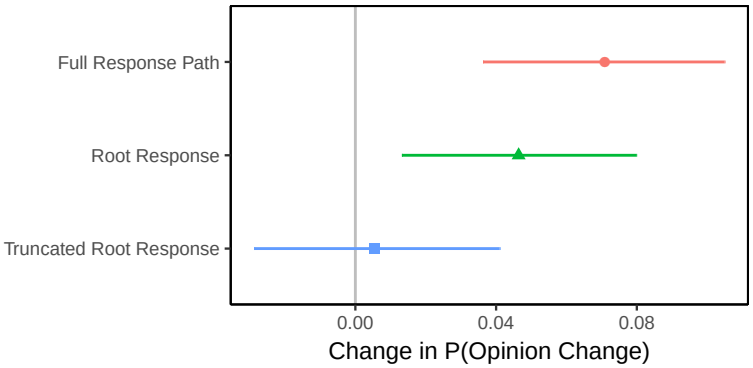


Figure 7: Moral congruence and persuasiveness: Change in predicted probability of opinion change ($\Delta$ awarded) when MFT congruence is increased from its minimum (no overlap in moral terms) to their empirical maximum (equal emphasis on each moral foundation) (including 95% confidence intervals). Estimates are based on logit models with standard errors clustered by discussion thread. Full model results are displayed in the appendix, Table E.2.

irrespective of whether I examine the content of the entire discussion (full response path), or restrict the analysis to each user's first post challenging the OP (root response). However, the mean difference in moral congruence does not reach conventional levels of statistical significance after truncating root responses to the same length for each pair. This finding could either suggest that the truncation procedure introduces too much noise to recover any differences, or alternatively that the measure of moral congruence is confounded by the differential length of successful and unsuccessful posts. The former seems less likely to be an issue because I recovered only marginal differences in raw dictionary term percentages between raw and truncated rooted responses in the previous section.

It is also important to emphasize that the higher moral congruence among persuasive posts is by no means driven by the fact that successful arguments use similar language to the opening statement *in general*. Quite contrary, Tan et al. (2016, 618) concluded in their study that when looking at the entire vocabulary of responses (excluding stopwords), then persuasive arguments used significantly more *different wording* than original post. In other words, a similar general vocabulary across all words is less persuasive, whereas a similar use of terms belonging to each moral foundation proved to be more persuasive. As such, the results presented here appear to

capture the unique persuasive effect of morally congruent arguments.

# 5 Conclusion

Political elites on both sides of the aisle routinely rely on moral rhetoric in order to bolster their views, which induces strong emotional reactions among citizens (Lipsitz, 2018) and can ultimately influence their attitudes (e.g., Clifford and Jerit, 2013; Clifford et al., 2015). As such, it does not seem surprising that the increasingly partisan and polarized environment in the United States has been linked to stronger tendencies among citizens to moralize politics (Garrett and Bankert, 2018). Is the only solution to overcome this trend to de-emphasize moral convictions when discussing political issues? Or is it rather the case that morality may even be helpful in overcoming disagreements as long as people rely on the same moral frameworks?

The present paper addresses these questions by contrasting two strands of research in moral psychology that lead to diverging predictions regarding the role of morality in political compromise. Previous work on moral conviction suggests that individuals who moralize politics should be less willing to compromise and therefore resist persuasion through moral appeals. On the other hand, moral foundations theory posits that compromise is indeed possible as long as the discussants use the same moral language.

Both competing hypotheses are tested by relying on a unique dataset of online discussions on the Reddit community CMV compiled by Tan et al. (2016). Overall, the empirical patterns support moral foundations theory and stand in contrast to predictions rooted in the literature on moral conviction. While general levels of moralization have little impact on argument persuasiveness, the results show that an argument's moral congruence with the discussant's opening statement increases the likelihood of changing his or her view. As such, moral appeals can facilitate compromise and change people's minds as long as they are consistent with their existing moral frameworks. Rather than automatically driving people further apart, moral appeals might therefore help bridge the growing divide between liberals and conservatives. More broadly, the

paper shows that the field of moral psychology stands to benefit from a further integration of two prominent theoretical frameworks that developed largely independent of each other and—unfortunately—still exhibit relatively little interconnections.

At the same time, the analysis presented here has important limitations. One of the biggest potential issues is the fact that the matched argument pairs differ in length, which may confound the relationship between morality and persuasiveness. I addressed this concern by only examining measures that are standardized by the total number of words in each post and by examining root responses in addition to full response paths. The results are largely robust to these varying specifications, with the important exception of the effect of moral congruence in the truncated root response. More generally, while it is a substantial advantage that the discussions on CMV cover a wide range of topics, it can be argued that some of them are ultimately irrelevant for moral considerations (such as software and technology). On the other hand, such inherently non-political and non-moral discussions should not induce any systematic biases between successful and non-successful arguments. Indeed, excluding discussions that focus on non-political issues does not change the substantive results of the analysis presented here (see Appendix F). I leave it to future research to leverage more controlled environments and focus on specific (political) issues—for example in the context of laboratory experiments. In contrast to framing studies conducted in the past, however, it is time to open the black box of conversations and directly examine the content of discussions in order to better understand the mechanisms underlying attitude change, persuasion, and compromise.

# References


Abramowitz, Alan I, and Kyle L Saunders. 2008. "Is polarization a myth?" *The Journal of Politics* 70 (02): 542–555.

Ahn, Toh-Kyeong, Robert Huckfeldt, and John Barry Ryan. 2010. "Communication, influence, and informational asymmetries among voters." *Political Psychology* 31 (5): 763–787.

Barabas, Jason. 2004. "How deliberation affects policy opinions." *American Political Science Review* 98 (04): 687–701.

Blei, David M, Andrew Y Ng, and Michael I Jordan. 2003. "Latent dirichlet allocation." *the Journal of machine Learning research* 3: 993–1022.

Chaiken, S., and A. H. Eagly. 1989. "Heuristic and Systematic Information Processing Within and Beyond the Persuasion Context." In *Unintended Thought*, ed. J. S. Uleman and J. A. Bargh. Vol. 212. New York: Guilford Press.

Clifford, Scott, and Jennifer Jerit. 2013. "How Words Do the Work of Politics: Moral Foundations Theory and the Debate over Stem Cell Research." *Journal of Politics* 75 (3): 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* 32 (2): 229–248.

Cobb, Michael D, and James H Kuklinski. 1997. "Changing minds: Political arguments and political persuasion." *American Journal of Political Science* 41 (1): 88–121.

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.

Druckman, James N, and Kjersten R Nelson. 2003. "Framing and deliberation: How citizens' conversations limit elite influence." *American Journal of Political Science* 47 (4): 729–745.

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.

Feinberg, Matthew, and Robb Willer. 2013. "The moral roots of environmental attitudes." *Psychological Science* 24 (1): 56–62.

Feinberg, Matthew, and Robb Willer. 2015. "From Gulf to Bridge When Do Moral Arguments Facilitate Political Influence?" *Personality and Social Psychology Bulletin* 41 (12): 1665–1681.

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.

Garrett, Kristin N, and Alexa Bankert. 2018. "The Moral Roots of Partisan Division: How Moral Conviction Heightens Affective Polarization." *British Journal of Political Science*: 1–20.

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. 2007. "The new synthesis in moral psychology." *science* 316 (5827): 998–1002.

Haidt, Jonathan. 2012. *The Righteous Mind: Why Good People are Divided by Politics and Religion*. New York: Random House.

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.

Hetherington, Marc J. 2001. "Resurgent mass partisanship: The role of elite polarization." *American Political Science Review* 95 (03): 619–631.

Hidey, Christopher, Elena Musi, Alyssa Hwang, Smaranda Muresan, and Kathy McKeown. 2017. "Analyzing the Semantic Types of Claims and Premises in an Online Persuasive Forum." In *Proceedings of the 4th Workshop on Argument Mining*. pp. 11–21.

Huckfeldt, Robert, Paul Allen Beck, Russell J Dalton, and Jeffrey Levine. 1995. "Political environments, cohesive social groups, and the communication of public opinion." *American Journal of Political Science* 39: 1025–1054.

Huddy, Leonie, Lilliana Mason, and Lene Aarøe. 2015. "Expressive Partisanship: Campaign Involvement, Political Emotion, and Partisan Identity." *American Political Science Review* 109 (01): 1–17.

Iyengar, Shanto, Gaurav Sood, and Yphtach Lelkes. 2012. "Affect, not ideology. A social identity perspective on polarization." *Public Opinion Quarterly* 76 (3): 405–431.

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.

Jhaver, Shagun, Pranil Vora, and Amy Bruckman. 2017. Designing for Civil Conversations: Lessons Learned from ChangeMyView. Technical report Georgia Institute of Technology.

Karpowitz, Christopher F, Tali Mendelberg, and Lee Shaker. 2012. "Gender inequality in deliberative participation." *American Political Science Review* 106 (03): 533–547.

Kertzer, Joshua D., Kathleen E. Powers, Brian C. Rathbun, and Ravi Iyer. 2014. "Moral Support: How Moral Values Shape Foreign Policy Attitudes." *Journal of Politics* 76 (3): 825–840.

Kidwell, Blair, Adam Farmer, and David M Hardesty. 2013. "Getting liberals and conservatives to go green: Political ideology and congruent appeals." *Journal of Consumer Research* 40 (2): 350–367.

Klar, Samara. 2014. "Partisanship in a social setting." *American Journal of Political Science* 58 (3): 687–704.

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.

Kraft, Patrick W. 2018. "Measuring Morality in Political Attitude Expression." *The Journal of Politics* 80 (3): 1028–1033.

Lazer, David, Brian Rubineau, Carol Chetkovich, Nancy Katz, and Michael Neblo. 2010. "The coevolution of networks and political attitudes." *Political Communication* 27 (3): 248–274.

Lipsitz, Keena. 2018. "Playing with Emotions: The Effect of Moral Appeals in Elite Rhetoric." *Political Behavior* 40 (1): 57–78.

Low, Michelle, and Ma Glenda Lopez Wui. 2015. "Moral Foundations and Attitudes Towards the Poor." *Current Psychology* 35: 650–656.

Manning, Christopher D., Prabhakar Raghavan, Hinrich Schütze et al. 2008. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

Mason, Lilliana. 2015. "I Disrespectfully Agree: The Differential Effects of Partisan Sorting on Behavioral and Issue Polarization." *American Journal of Political Science* 59 (1): 128–145.

Mendelberg, Tali, Christopher F Karpowitz, and J Baxter Oliphant. 2014. "Gender inequality in deliberation: Unpacking the black box of interaction." *Perspectives on Politics* 12 (01): 18–44.

Mutz, Diana C. 2002. "Cross-cutting social networks: Testing democratic theory in practice." *American Political Science Review* 96 (01): 111–126.

Nelson, Thomas E, and Jennifer Garst. 2005. "Values-based Political Messages and Persuasion: Relationships among Speaker, Recipient, and Evoked Values." *Political Psychology* 26 (4): 489–516.

Petty, Richard E, and John T Cacioppo. 1986a. *Communication and persuasion: Central and peripheral routes to attitude change*. New York: Springer.

Petty, Richard E, and John T Cacioppo. 1986b. "The elaboration likelihood model of persuasion." *Advances in experimental social psychology* 19: 123–205.

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." *Journal of Politics* 76 (2): 380–397.

Ryan, Timothy J. 2017. "No compromise: Political consequences of moralized attitudes." *American Journal of Political Science* 61 (2): 409–423.

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): 895–917.

Skitka, Linda J, and G Scott Morgan. 2014. "The social and political implications of moral conviction." *Political Psychology* 35 (S1): 95–110.

Tan, Chenhao, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. "Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions." In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee pp. 613–624.

Wei, Zhongyu, Yang Liu, and Yi Li. 2016. "Is this post persuasive? Ranking argumentative comments in online forum." In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. Vol. 2 pp. 195–200.

# Appendices

*Online Appendix:*
**Change My View – Do Moral Appeals Facilitate Compromise?**

# Appendix A Basic Rules on the Subreddit ChangeMyView

Below is a summary of the set of rules to participate in discussions on /r/ChangeMyView as described in April 2018. The current rules can be viewed at https://www.reddit.com/r/changemyview/wiki/rules

**Rules for submission of new discussion posts:**

A Explain the reasoning behind your view, not just what that view is (500+ characters required).

B You must personally hold the view and demonstrate that you are open to it changing.

C Submission titles must adequately sum up your view and include "CMV:" at the beginning.

D Posts cannot express a neutral stance, carry a risk of personal endangerment, be self-promotional, or discuss this subreddit (visit r/ideasforcmv instead).

E Only post if you are willing to have a conversation with those who reply to you, and are available to start doing so within 3 hours of posting.

**Rules for commenting in existing discussions:**

1 Direct responses to a CMV post must challenge at least one aspect of OPs stated view (however minor), or ask a clarifying question.

2 Don't be rude or hostile to other users.

3 Refrain from accusing OP or anyone else of being unwilling to change their view.

4 Award a delta if you've acknowledged a change in your view. Do not use deltas for any other purpose.

5 Comments must contribute meaningfully to the conversation.

# Appendix B Moral Foundations Dictionary

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

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

### Care

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

### Fairness

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

### Loyalty

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

### Authority

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

### Sanctity

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

### General Morality

bad, blameless, canon, character, commendable, correct, decen*, doctrine, ethic*, evil, exemplary, good, goodness, honest*, ideal*, immoral*, indecen*, integrity, laudable, lawful*, legal*, lesson, moral*, noble, offend*, offensive*, piety, pious, praiseworthy, principle*, proper, righteous*, transgress*, upright, upstanding, value*, wholesome*, wicked*, worth*, wretched*, wrong*

# Appendix C Distribution of Moral Foundation Proportions in Paired Data

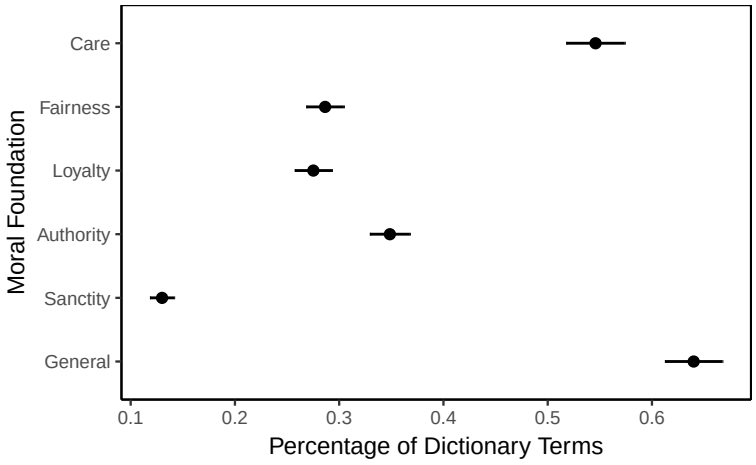


Figure C.1: Moral Foundations in Paired Data: Average percentage of dictionary terms relative to the total number of words in each original post starting a discussion (including 95% confidence intervals). Compared to the figure in the main text, this plot only includes opening statements that are part of the matched pair selection to analyze persuasive arguments.

# Appendix D Structural Topic Model Results

## I Original Posts

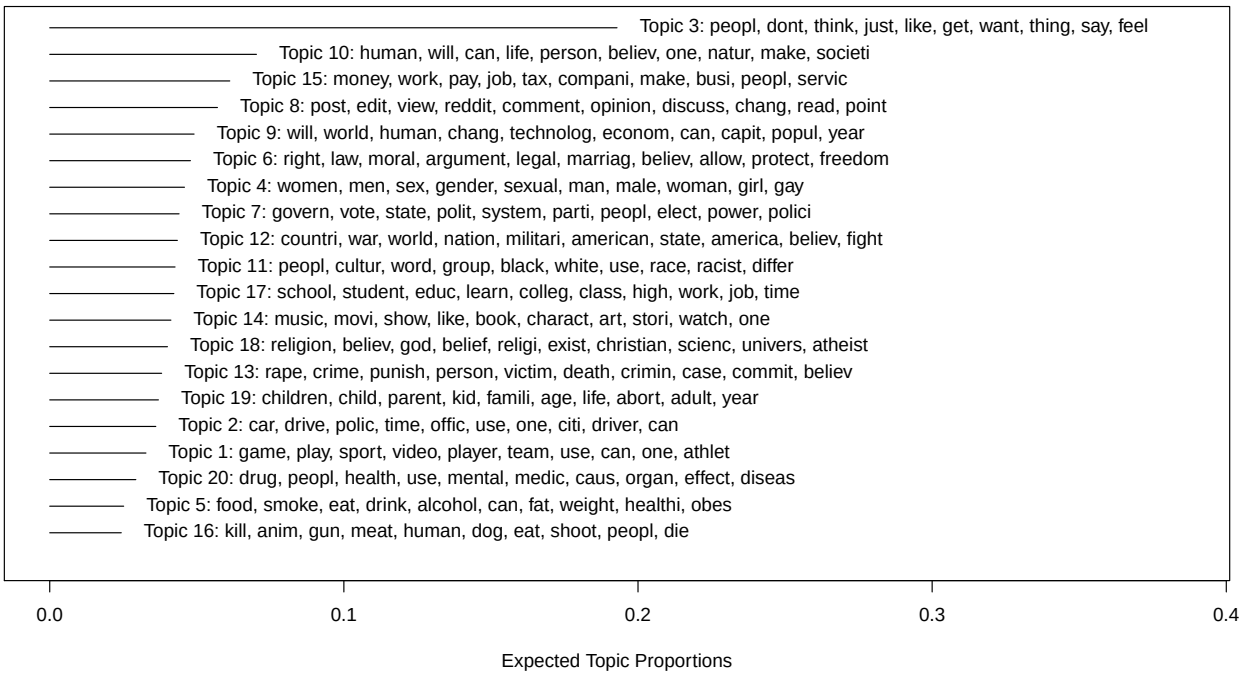


Figure D.1: Average topic proportions in opening statements on `/r/ChangeMyView/` based on a structural topic model with 20 topics (c.f., Roberts et al., 2014). The plot additionally displays the ten most likely terms associated with each respective topic.

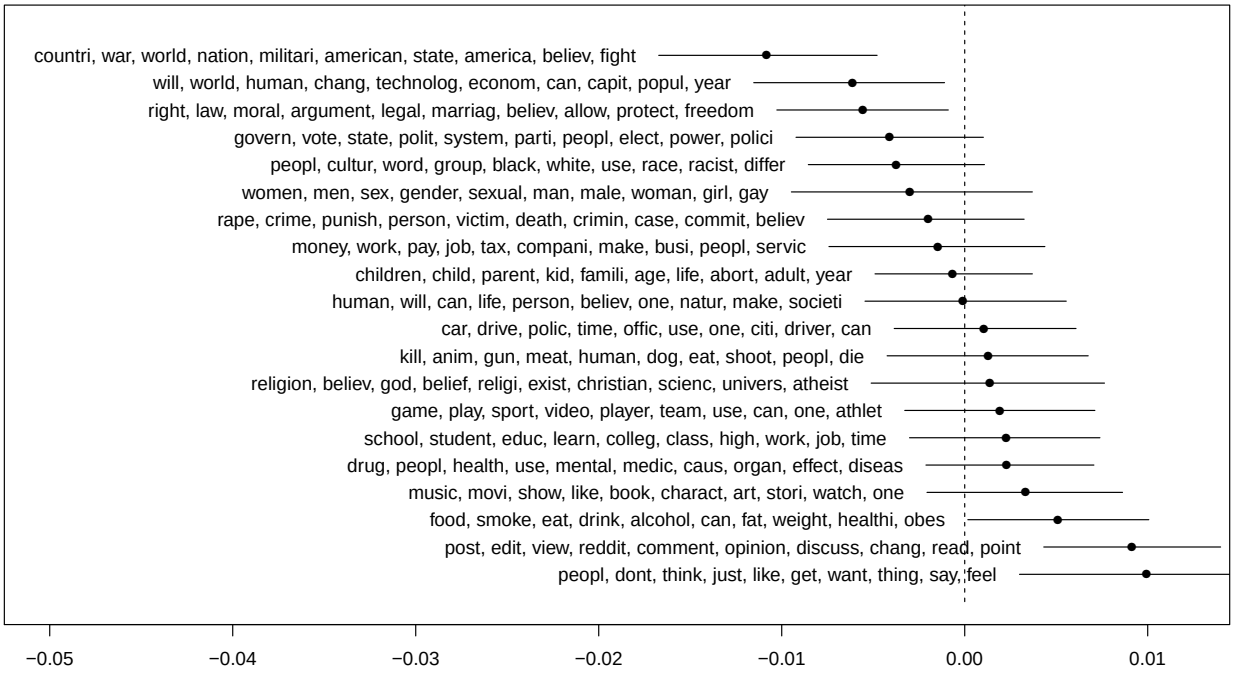


Figure D.2: Differences in topic proportions between opening statements on `/r/ChangeMyView/` that resulted in opinion change ($\Delta$ awarded) versus not (including 95% confidence intervals). Estimates are based on the structural topic model described in the previous figure. Positive values indicate higher topic prevalence among discussions that resulted in opinion change and vice versa. Labels are based on the ten highest probability terms related to the topic.

# II Responses Challenging the OP

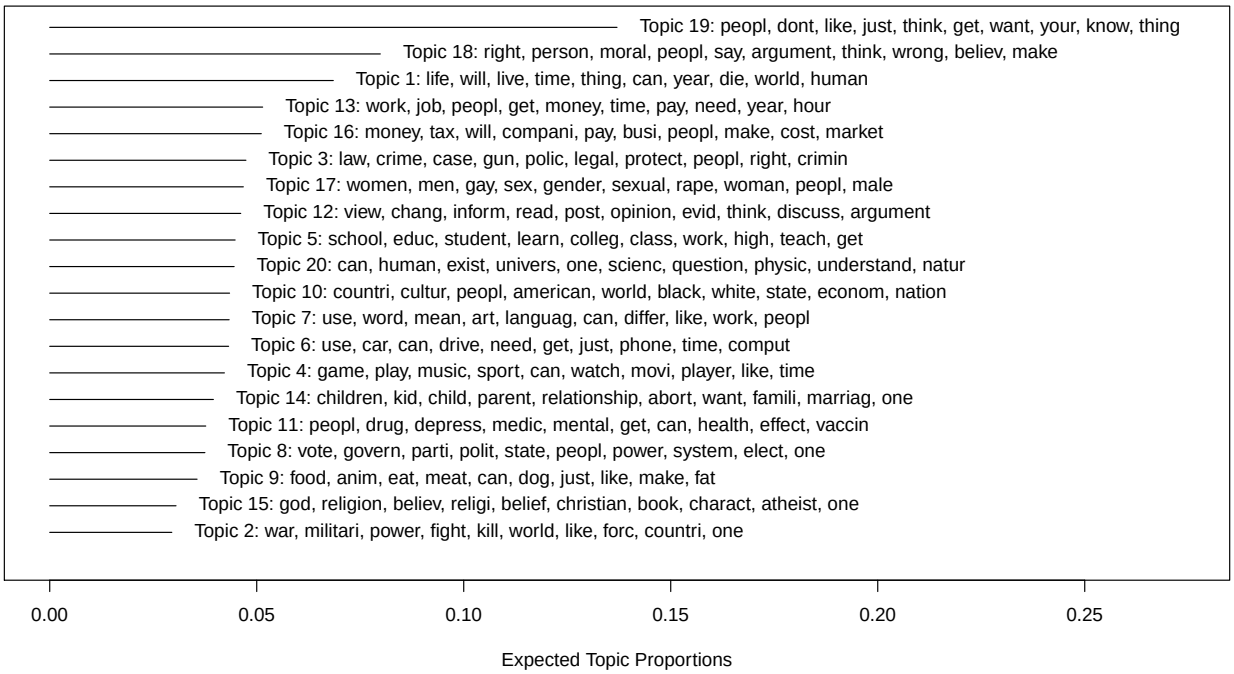


Figure D.3: Average topic proportions in posts challenging the OP on `/r/ChangeMyView/` based on a structural topic model with 20 topics (c.f., Roberts et al., 2014). The plot additionally displays the ten most likely terms associated with each respective topic.

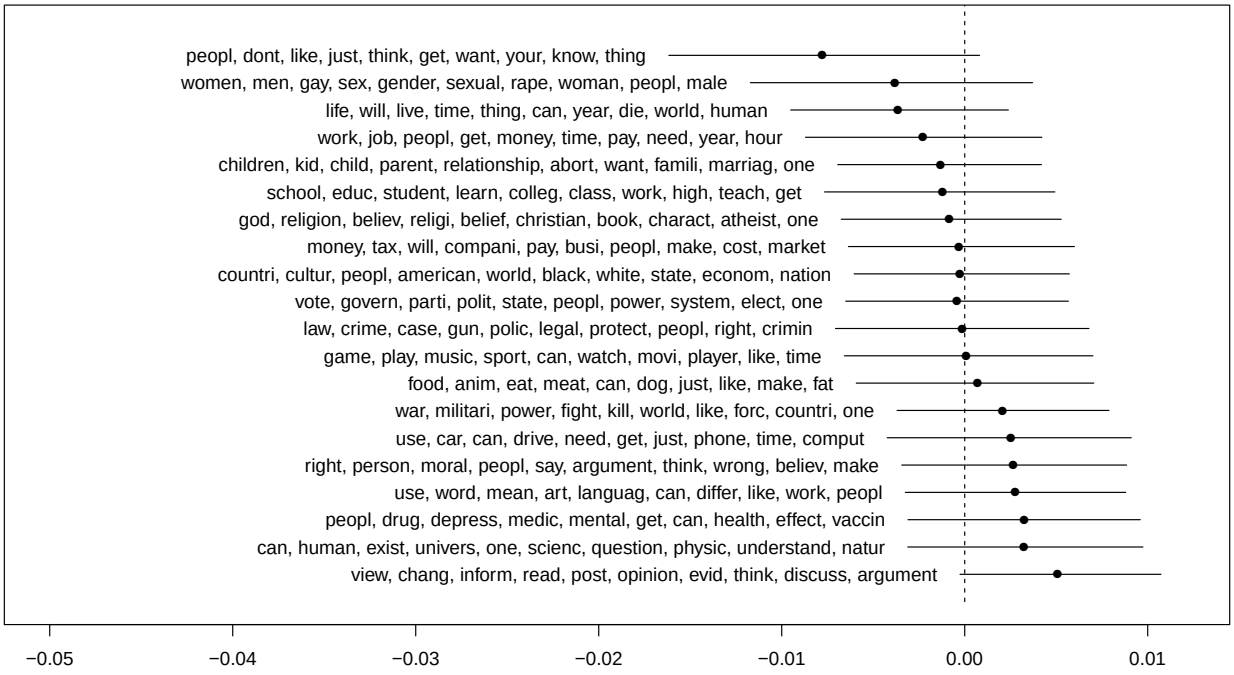


Figure D.4: Differences in topic proportions between persuasive and non-persuasive responses challenging the OP (including 95% confidence intervals). Estimates are based on the structural topic model described in the previous figure. Positive values indicate higher topic prevalence among posts that received a $\Delta$ by the OP and vice versa. Labels are based on the ten highest probability terms related to the topic.

# Appendix E Tables of Model Estimates

Table E.1: Logit models predicting argument persuasiveness as a function of moral word use (measured via MFT dictionary proportions). Positive coefficients indicate higher probability of changing the OPs' mind ($\Delta$ awarded). Standard errors (clustered by discussion thread) in parentheses. Estimates are used for Figure 6 in the main text.

| Variable | Full Response Path | Root Response | Truncated Root Response |
|---|---|---|---|
| Care | -0.004 | -0.009 | -0.007 |
| | (0.025) | (0.023) | (0.023) |
| Fairness | -0.029 | -0.024 | -0.032 |
| | (0.036) | (0.033) | (0.031) |
| Loyalty | 0.005 | 0.017 | 0.018 |
| | (0.035) | (0.033) | (0.03) |
| Authority | 0.003 | -0.005 | 0.009 |
| | (0.03) | (0.028) | (0.027) |
| Sanctity | -0.033 | -0.005 | -0.022 |
| | (0.047) | (0.046) | (0.044) |
| General | -0.010 | -0.010 | -0.004 |
| | (0.024) | (0.023) | (0.022) |
| Intercept | 0.018 | 0.015 | 0.009 |
| | (0.024) | (0.023) | (0.022) |
| N | 6304 | 6304 | 6304 |
| Log-Likelihood | -4369 | -4369 | -4369 |

Table E.2: Logit models predicting argument persuasiveness as a function of moral congruence with OPs' opening statements (measured via cosine similarity in MFT dictionary results). Positive coefficients indicate higher probability of changing the OPs' mind ($\Delta$ awarded). Standard errors (clustered by discussion thread) in parentheses. Estimates are used for Figure 7 in the main text.

| Variable | Full Response Path | Root Response | Truncated Root Response |
|---|---|---|---|
| Moral Congruence | 0.290 | 0.188 | 0.019 |
| | (0.056) | (0.056) | (0.054) |
| Intercept | -0.147 | -0.092 | -0.008 |
| | (0.028) | (0.027) | (0.024) |
| N | 6304 | 6304 | 6304 |
| Log-Likelihood | -4361 | -4366 | -4370 |

# Appendix F Excluding Non-political Discussions

For the purpose of this paper, I decided against filtering out subsets of discussions based on their thematic (or political) relevance since such a strategy could raise additional concerns about potential selection bias. However, it is worth exploring whether the substantive results are robust when focusing only on discussions revolving around politics. In the following, I therefore replicate all major analyses of the paper after excluding discussions related to distinctly non-political topics (e.g., "Food" or "Entertainment"). The following figure shows the total number of discussions that focus on political versus non-political issues (based on the LDA results).

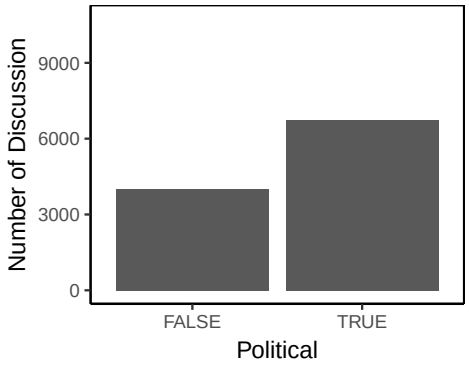


Figure F.1: Number of discussions on `/r/ChangeMyView/` that focus on political versus non-political issues (based on the LDA results).

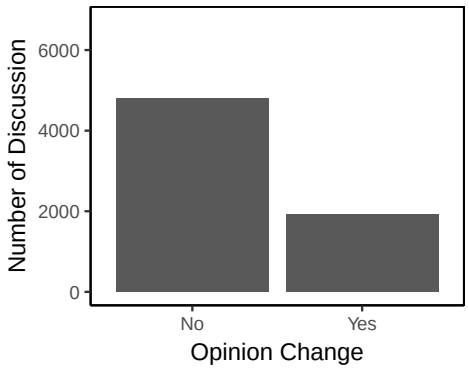


Figure F.2: Political discussions only – Number of discussions on `/r/ChangeMyView/` that resulted in opinion change (at least one $\Delta$ awarded by OP) versus not.

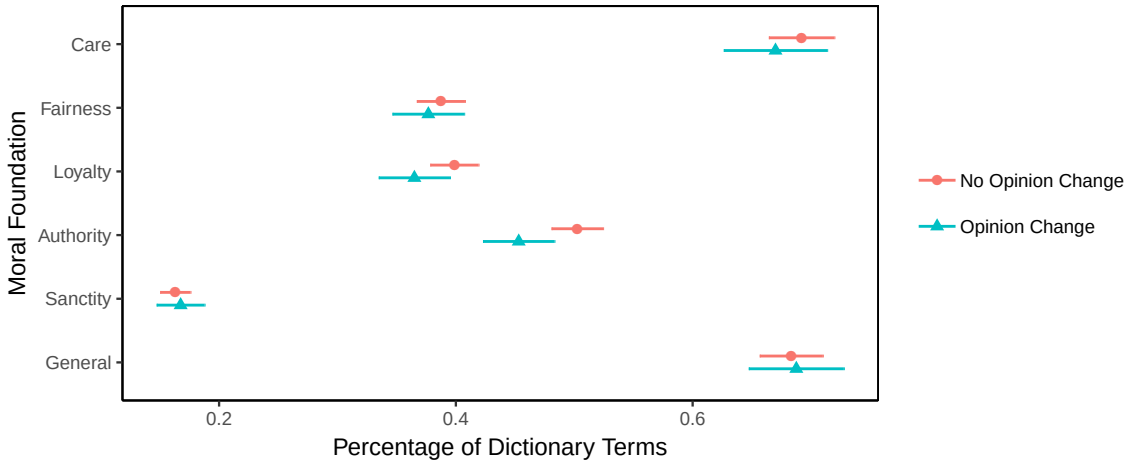


Figure F.3: Political discussions only – Moral foundations and persuadability: Average percentage of dictionary terms relative to the total number of words in each opening statement beginning a discussion, comparing discussions where the OP subsequently awarded a $\Delta$ (opinion change) or not (including 95% confidence intervals).

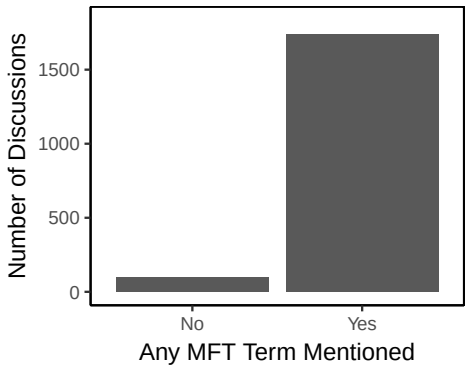


Figure F.4: Political discussions only – Number of opening statements in the paired argument data that included *any* term mentioned in the MFT dictionary.

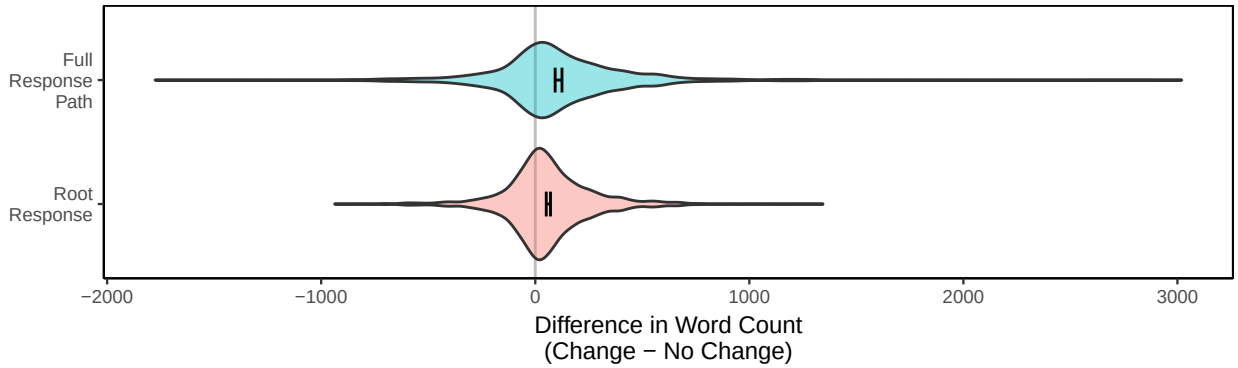


Figure F.5: Political discussions only – Difference in response lengths between successful and unsuccessful counterarguments. The narrow black bars display the 95% confidence interval of mean differences.

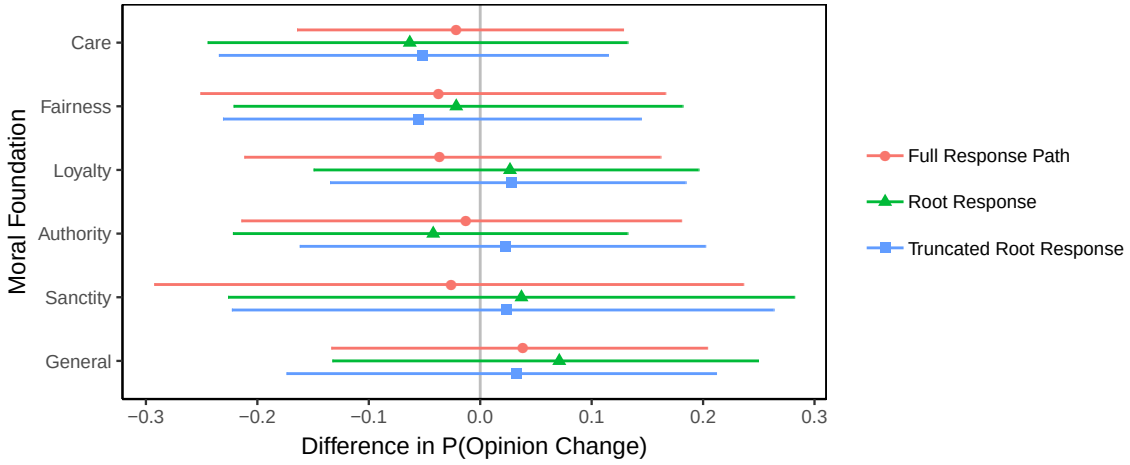


Figure F.6: Political discussions only – Moral foundations and persuasiveness: Change in predicted probability of opinion change ($\Delta$ awarded) when MFT dictionary term proportions are increased from their minimum (no moral terms mentioned) to their empirical maximum, holding the remaining foundations constant at their mean (including 95% confidence intervals). Estimates are based on logit models with standard errors clustered by discussion thread.

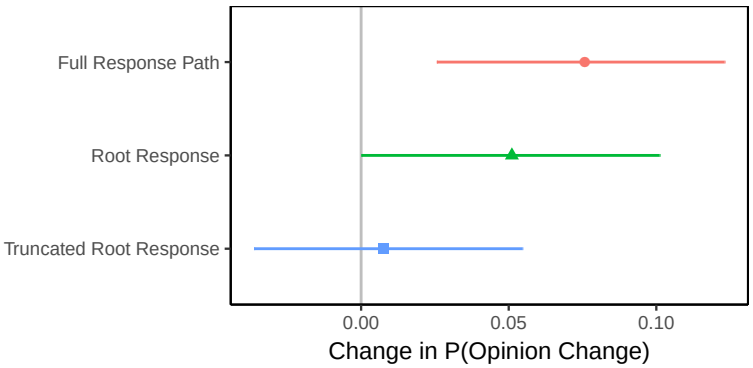


Figure F.7: Political discussions only – Moral congruence and persuasiveness: Change in predicted probability of opinion change ($\Delta$ awarded) when MFT congruence is increased from its minimum (no overlap in moral terms) to their empirical maximum (equal emphasis on each moral foundation) (including 95% confidence intervals). Estimates are based on logit models with standard errors clustered by discussion thread.