

# 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PLEASE DO NOT CITE OR REDISTRIBUTE WITHOUT PERMISSION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

---

\*Ph.D. Candidate, Stony Brook University, [patrick.kraft@stonybrook.edu](mailto:patrick.kraft@stonybrook.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, ideological divides may be overcome if people speak the same "moral language." 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, however, 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) demonstrates 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. 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<sup>1</sup> (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 results show that moral arguments can facilitate compromise, but only to the extent that they are congruent with the moral framework of the opposing discussant.

---

<sup>1</sup><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 updated 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). From this distinction, it follows 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.

As a consequence, there is no general answer to the question whether complex or simple arguments are more effective in changing people's views. For example, Cobb and Kuklinski (1997) analyze the persuasive effect of different argument types 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. One explanation for these 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 which they are more likely to rely on peripheral cues which renders complex arguments ineffective. However, motivation to engage in central processing is not only driven by individual predispositions and the nature of the underlying issue itself. 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.

Similarly, moral appeals may influence the effectiveness of persuasive communications through multiple ways. 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 affect the degree to which the message is scrutinized with more elaboration (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 provide 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 showed 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 as a distinct characteristic of attitude intensity 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 ultimately suggests that counterarguments that emphasize the 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 (SDO) and right-wing authoritarianism (RWA) (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.”

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, moral foundations theory contends that agreement can be facilitated if two discussants focus on the same underlying moral dimensions. The underlying assumption that emphasizing the same foundations can facilitate compromise has important implications—especially in our current political environment. Somewhat surprisingly, however, this claim has rarely been subjected to a direct empirical test.

## 1.3 Hypotheses

The structure and dynamics of political discussions can be very complex, which makes 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 example of a conversation between two discussants. Person  $\mathcal{A}$  is set out to defend a specific opinion and person  $\mathcal{B}$  wants to challenge  $\mathcal{A}$ 's view. Suppose that  $\mathcal{A}$  makes an opening statement justifying her opinion using moral arguments. Person  $\mathcal{B}$  may then engage in the discussion and try to persuade  $\mathcal{A}$  using either moralized or non-moralized counterarguments. Both theoretical perspectives described in the previous section suggest contrasting hypotheses regarding the persuasiveness of  $\mathcal{B}$ 's appeals:

*H1 (Moral Conviction):* Moralized counterarguments are going to be *less* persuasive than non-moralized counterarguments.

*H2 (Moral Foundations):* Moralized counterarguments that are congruent with the opening statement's moral framework are going to be *more* persuasive than non-moralized or morally incongruent counterarguments.

In this directional model of a discussion, only  $\mathcal{A}$  stands to maintain or change her view, whereas  $\mathcal{B}$  only attempts to persuade her discussant. Compromise is achieved in this scenario if  $\mathcal{B}$  is able to persuade  $\mathcal{A}$  to change her attitude. Unfortunately, even in this simplified context, it is hard to test the predictions using simple framing experiments employed in previous studies. However, the next section will illustrate how conversations on the Reddit community `/r/ChangeMyView` provide an ideal environment to test both competing hypotheses.

## 2 The Subreddit “ChangeMyView”

Users can initiate a discussion on CMV 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. The community is dedicated to civil discourse and encourages OPs to be open to changing their views and to identify convincing counterarguments genuinely (e.g., Jhaver, Vora, and Bruckman, 2017).<sup>2</sup> This online discussion format provides an ideal environment to explore the correlates of argument persuasiveness. In contrast to past framing studies, it allows for an examination of a wide array of issues. More importantly, 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 rather than source cues and identity-related factors. Lastly, the fact that OPs evaluate a multitude of available counterarguments allows for a unique counterfactual design to study argument persuasiveness.

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). For the analyses presented here, I rely on a dataset of matched argument pairs extracted from CMV by Tan et al. (2016), who studied the effects of discussion dynamics and linguistic features on persuasion. My analysis extends previous results by emphasizing the role of moral content in attitude change.

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 Reddit community where participants begin a discussion thread by stating their personal opinion on a chosen issue. Other users are invited to challenge the original posters' (OP) arguments and try to persuade them otherwise. The OPs then respond to the arguments brought forward and explicitly identify forum entries that changed their view by awarding a "Delta" ( $\Delta$ ). The participation rules of the community foster civil exchange of arguments—even for divisive issues—and OPs are encouraged to award  $\Delta$ s genuinely.<sup>3</sup> To date, the subreddit has more than 500,000 subscribed

---

<sup>2</sup>See below for more detailed information on the rules and mechanics of discussions on CMV.

<sup>3</sup>The current set of rules for original posts as well as responses can be accessed at <https://www.reddit.com/r/changemyview/wiki/rules>. See Appendix for an overview.

users.

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

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 he awarded a  $\Delta$ . Ultimately, these structured online discussions on CMV provide an ideal opportunity to study the persuasiveness of moral appeals. They 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 found to be persuasive enough to change their views.

Tan et al. (2016) make use of this structured discussion and explore interaction dynamics on CMV by analyzing linguistic features 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. The analyses presented below leverage the dataset compiled by Tan et al. (2016). In this context, 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 characteristics (i.e., *how it is expressed*) to

predict persuasiveness. The following analyses explicitly turns to the effects of moral content on discussion outcomes.

Before turning to the analysis, I want to provide a brief overview of the range of topics that are discussed on CMV. Based on the set of more than 10,000 original discussion posts included in the data, I estimated a basic topic model extracting clusters of co-occurring terms via Latent Dirichlet Allocation [Blei, Ng, and Jordan \(2003\)](#). The topic models are based on the set of original opening statements (disregarding subsequent comments), which 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 displays the average topic proportions of opening statements based on the model.

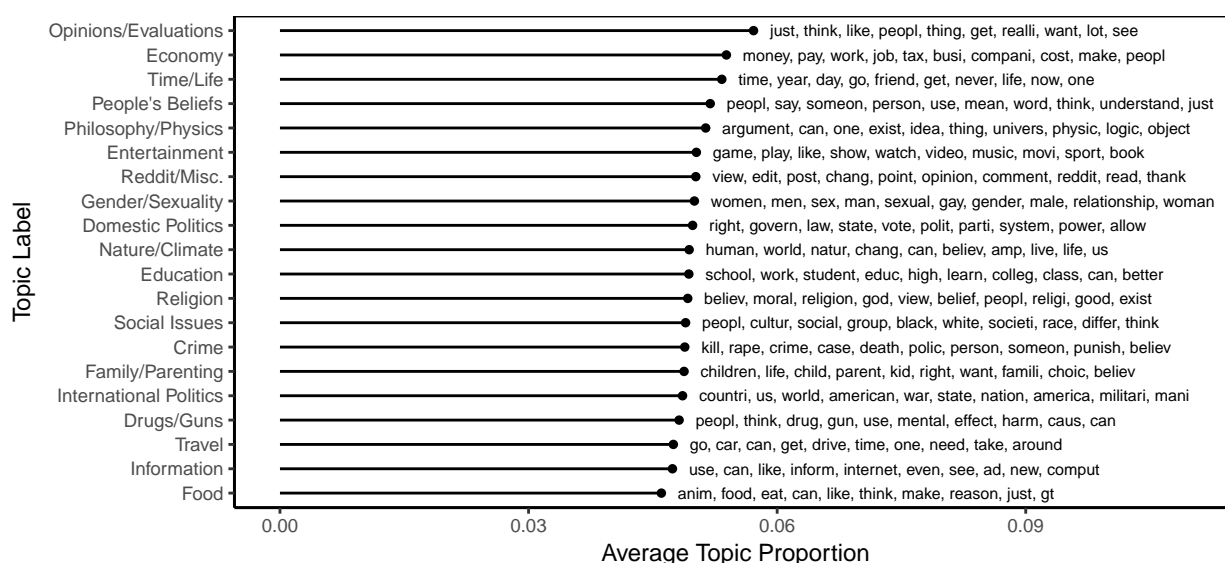


Figure 1: Average topic proportions in original posts on /r/ChangeMyView/ based on a basic LDA model with 20 topics. The plot additionally displays the ten most likely terms associated with each respective topic.

Discussions on CMV range across a variety of topics such as economic issues, gender/sexuality, or domestic and international politics. Additional analyses included in the appendix show that there are slight differences in the extent to which topics differ in the likelihood of persuasion in discussions, although these differences are surprisingly small. The following sections turn to the role of moral appeals in facilitating or inhibiting compromise.

### 3 Opinion Change in Online Discussions

As an initial step, I examine the extent to which OPs are willing to award  $\Delta$ s during a discussion in the first place. Figure 2 displays the number of discussions included in the dataset where OPs indicated that one of the responses changed their mind.

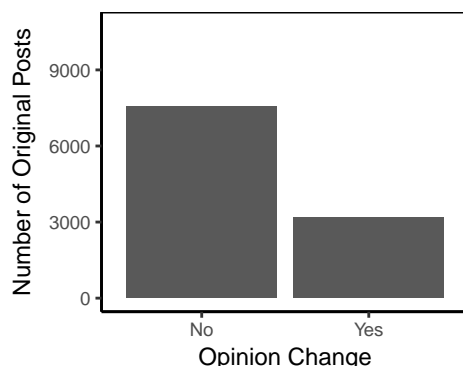


Figure 2: Number of original posts on /r/ChangeMyView/ that resulted in opinion change ( $\Delta$  awarded by author) versus not.

In about two thirds of the 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 discussions where OPs indicated that at least one of the responses changed their views.

In their original study, Tan et al. (2016) investigated linguistic and style patterns such as the use of personal pronouns as well as formatting that predicted resistance to persuasion among OPs. They conclude that “comparative adjectives and adverbs are a sign of malleability, while superlative adjectives suggest stubbornness.” The goal of the present analysis, however, is to go beyond pure linguistic style and explore the role of moral content in persuasion. To capture moral appeals in discussions, I rely on the the Moral Foundations 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. The full dictionary is included in Appendix A.

Figure 3 displays the percentage of dictionary terms belonging to each moral foundation relative to the total number of words comparing in opening statements beginning a discussion. Most

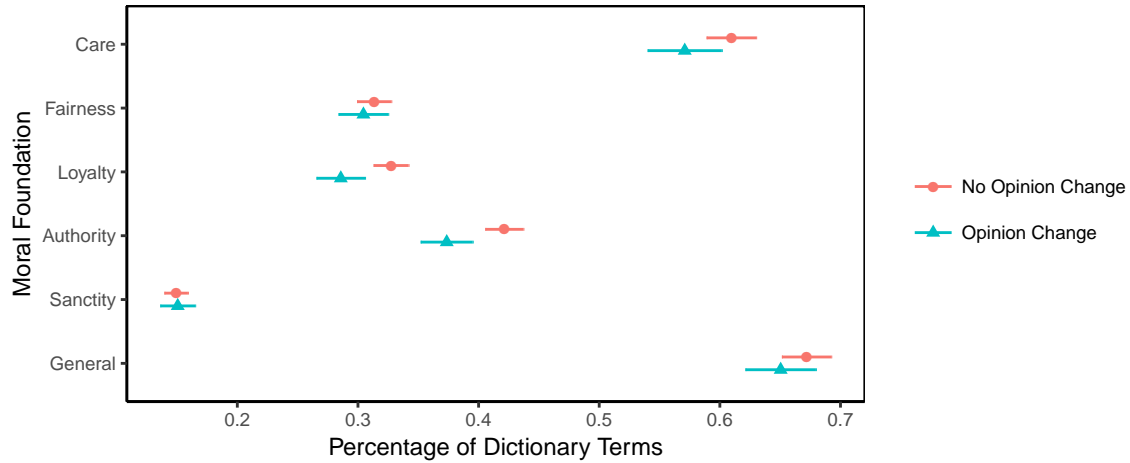


Figure 3: Moral Foundations and Persuadability: Average percentage of dictionary terms relative to the total number of words in each original post starting a discussion, comparing original posts where the author subsequently awarded a  $\Delta$  (opinion change) or not (including 95% confidence intervals).

importantly, the plot compares the reliance on moral terms between OPs that later reported that they changed their view versus OPs who did not. Interestingly, the distribution of dictionary terms across foundations is strikingly similar to the average proportions of moral terms in open-ended responses in the American National Election study reported in Kraft (2018). More importantly, the percentage of dictionary terms among opening statements that resulted in opinion change are significantly larger for two foundations, namely authority and loyalty ( $p < .01$  in both cases after accounting for multiple comparisons using Bonferroni correction). 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. We obtain a similar result after aggregating all moral dictionary terms in one category: OPs who not persuadable on CMV make stronger use of moral terms *in general* than OPs who indicate that the discussion changed their view ( $p < .001$ ). This basic result is consistent with the moral conviction literature, which posits that people who hold moralized attitudes are less willing to compromise (e.g., Skitka, Bauman, and Sargis, 2005; Ryan, 2014, 2017).

## 4 What Makes an Argument Persuasive?

However, the main interest of this paper is not the persuadability of OPs (which is certainly driven by unobserved confounding variables). Instead, we want to look at the persuasiveness of counterarguments that are made in response to the opening statements.

The previous section demonstrated that the usage of moral terms in initial statements by OPs is predictive of the malleability of their opinions. Next, I directly examine the effectiveness of specific counterarguments that were brought forward to challenge the OPs' views. Tan et al. (2016) propose a selection of argument pairs that respond to an original post with only one being successful in changing the OPs view while being relatively similar in terms of their word choice.

[Elaborate on argument pair selection etc.]

More specifically, Tan et al. (2016) select similar argument pairs based on their Jaccard similarity:

$$\text{Jaccard}(A, B) = \frac{|A \cap B|}{|A \cup B|}, \quad (1)$$

where  $A$  and  $B$  are the sets of words in both counterarguments. In other words, they match each successful counterargument to an unsuccessful response that shares the largest proportion of common words (disregarding stopwords).<sup>4</sup> As Tan et al. (2016, 5) describe: "This leads to a balanced binary prediction task: which of the two lexically similar rooted path-units is the successful one?"

The analyses reported below rely on the approach implemented by Tan et al. (2016) to select matched argument pairs. To reiterate, I therefore focus on discussions in which OPs awarded at least one  $\Delta$ . The response that received a  $\Delta$  is then matched to another counterargument that was not successful in changing the OPs view but is as similar as possible in terms of its vocabulary. Note that this strategy should make it difficult to find differences in the moral foundation dictionaries as argument pairs are matched based on their lexical similarity.

One might worry that this selection on discussions where OPs ultimately awarded a  $\Delta$  might

---

<sup>4</sup>Elaborate on additional selection criteria.

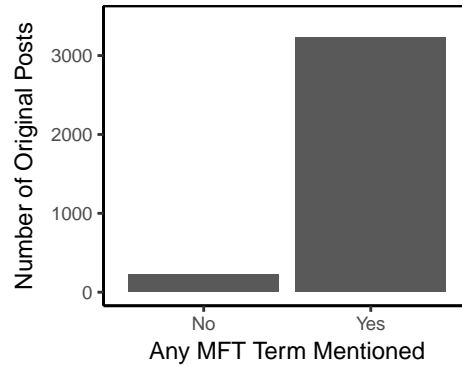


Figure 4: Number of original posts in the paired argument data that included *any* term mentioned in the moral foundation dictionary.

discard all cases where the initial statement emphasized moral considerations. Luckily, that is not the case. Figure 4 shows that almost all of the opening statement 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 (see Appendix B for details).

However, one issue that remains unresolved using this approach is that the matching procedure only focuses on the set of words that are used and does not take into account the length of each response. 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 be problematic when examining the effect of moral dictionary terms.

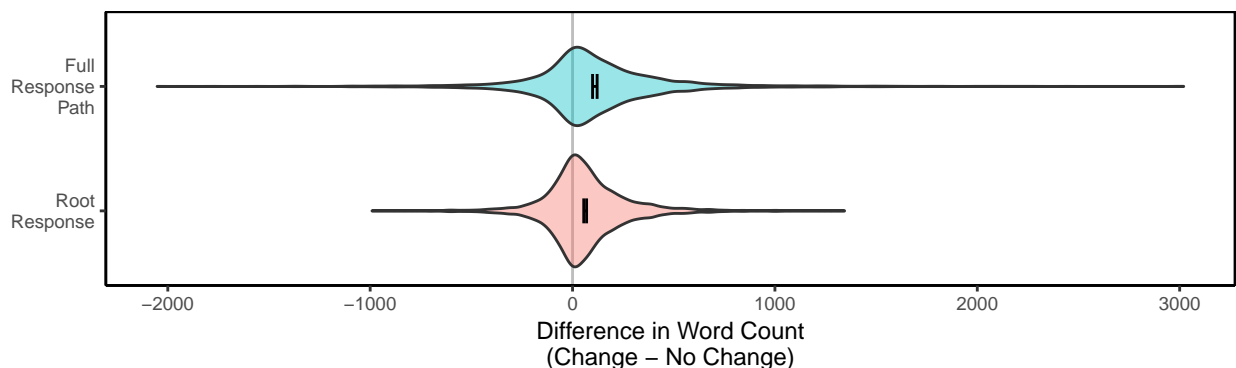


Figure 5: Difference in response lengths between successful and unsuccessful counterarguments. The narrow black bars display the 95% confidence interval of mean differences.



While it could be argued that the proportion of dictionary terms should be unaffected by differences in argument length, I take additional precautions proposed by Tan et al. (2016) to check the robustness of the results. First, I do not only examine differences when looking at the entire response path in the discussion between two users, but restrict the analysis on the initial response to the opening statement only. As can be seen in figure 5, the differences in word counts between argument pairs are significantly smaller. As a second step to account for the issue of response length, I additionally truncate the longer response of a given pair such that both responses contain the same number of words. Using this framework, I now turn to the analysis of the persuasiveness of moral arguments made within discussion on CMV.

#### 4.1 Moral Appeals are Futile...

Figure 6 displays the difference in dictionary term proportions between matched discussion contributions that were successful versus unsuccessful. Positive values indicate that arguments that received a  $\Delta$  contained a larger percentage of dictionary terms, and vice versa.

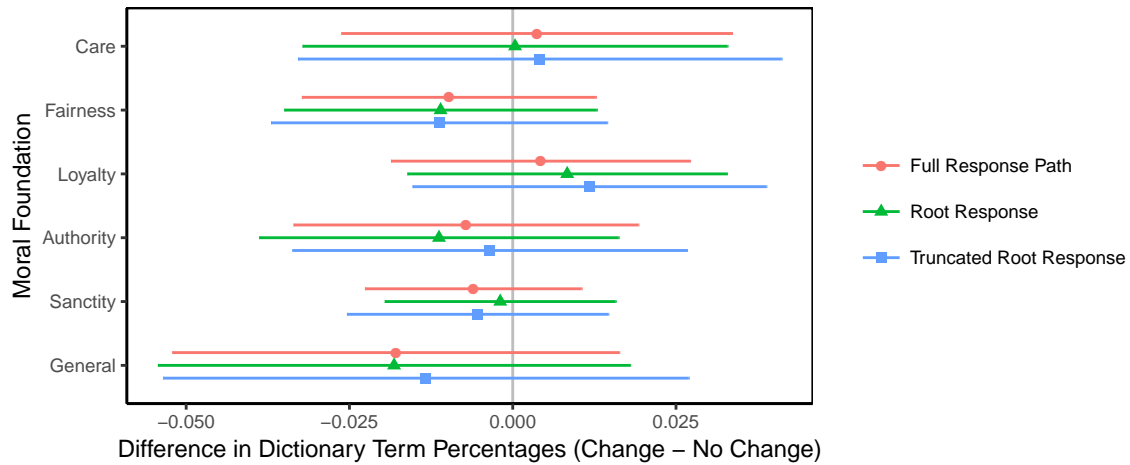


Figure 6: Moral Foundations and Persuasiveness: Average difference of dictionary term percentages relative to the total number of words in each post, comparing counterarguments where the OP subsequently awarded a  $\Delta$  (opinion change) or not (including 95% confidence intervals).

The results show that evoking moral considerations in counterarguments does not affect the likelihood of changing the OP's view on a given issue. This result is not consistent with the moral

conviction literature which would have predicted that moral appeals should in fact decrease the likelihood of compromise and opinion change.

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

However, moral foundations theory argues that we cannot view moral appeals independent of the moral framework of the discussion partner. Instead, what is seen as decisive is the congruence in moral arguments between both discussants. I measure moral congruence by computing the similarity in dictionary scores between opening statements by OPs and the respective counterarguments. More formally, moral congruence is written as:

$$\text{moral congruence} = \frac{\vec{x} \cdot \vec{y}}{||\vec{x}|| ||\vec{y}||}, \quad (2)$$

where  $\vec{x}$  is the vector of dictionary counts in the OP's opening statement and  $\vec{y}$  is the respective vector for a response. The measure ranges from 0 (no moral overlap) to 1 (equal emphasis on the same moral foundations).

Figure 7 displays the difference in moral congruence between successful and unsuccessful counterarguments. Positive values indicate higher congruence in moral appeals among counterarguments that were ultimately awarded a  $\Delta$  by OPs. The results clearly show that congruence in moral appeals is larger in persuasive than non-persuasive arguments.

This finding is especially noteworthy since the original analysis by Tan et al. (2016, 6) concluded that persuasive arguments used more different wording from the original post.

## 5 Conclusion

The results presented in this paper are consistent with moral foundations theory and contradict previous research on moral convictions. While the general moralization of arguments appeared to have had little impact on their persuasiveness, the results show that moral congruence with initial

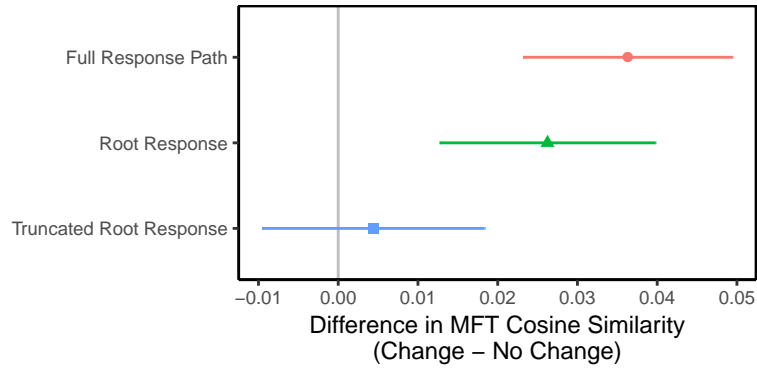


Figure 7: Moral congruence and persuasiveness: Average differences between successful and unsuccessful counterarguments in cosine similarities with the opening statements (including 95% confidence intervals).

statements are more likely to be effective in changing a person's view. As such, moral appeals can facilitate compromise and change people's minds as long as they are consistent with their existing moral frameworks. To summarize, being able to speak the same moral language might therefore help to bridge the growing divide between liberals and conservatives.

## 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.
- 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.
- 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.
- 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* (in press): 1–11.

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