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PLEASE DO NOT CITE OR REDISTRIBUTE WITHOUT PERMISSION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 agreement and consensus building through deliberation.

Keywords: Moral Foundations, Attitude Change, Persuasion, Compromise

---
*Ph.D. Candidate, Stony Brook University, patrick.kraft@stonybrook.edu.

1 Introduction

Outline of argument:

- Political discussions often revolve around underlying basic principles, especially in a polarized environment (e.g., Garrett and Bankert 2018).

- Moral conviction literature suggests that moralization of political attitudes should make compromise more difficult (Ryan 2016 and others).

- However, recent studies on moral frames (grounded in MFT) show that liberals/conservatives are more likely to be persuaded when arguments are congruent with their own moral framework (Koleva et al., 2012; Kidwell, Farmer, and Hardesty, 2013)

- Gap in the literature: few/no studies directly examined the role of moral arguments/appeals in discussions (i.e., going beyond simple framing scenarios).

- Make a convincing argument for why `/r/ChangeMyView` is a perfect case to study the effect of moral arguments in discussions and fill this gap. Describe basic idea of the subreddit. (Tan et al., 2016)

2 Morality and the Potential for Compromise

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—building on Moral Foundations Theory—attributes this divide (at least partially) to fundamental differences in moral frameworks that guide liberal and conservative thinking (c.f., Haidt, 2012). According to this perspective, 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 ultimately shape political attitudes, evaluations, and behavior.

An important implicit assumption 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." The underlying assumption that emphasizing the same foundations can facilitate compromise has important implications—especially in our current political environment. Somewhat surprisingly, however, it has never been subjected to a direct empirical test.

There are reasons to believe that compromise is still difficult—and potentially even further impeded—if individuals argue on the basis of the same set of moral foundations. Skitka, Bauman, and Sargis (2005), for example, provide a different theoretical perspective than MFT by conceptualizing moralization as a unique feature of attitude strength. 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, they combine the following attributes: they 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 (2016) 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).

Ultimately, both perspectives in moral psychology 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. Consider for example a discussion between person $\mathcal{A}$ who is set out to defend a specific opinion and person $\mathcal{B}$ who wants to challenge $\mathcal{A}$'s view and convince her otherwise. Suppose that $\mathcal{A}$ makes an opening statement justifying her opinion using moral

arguments. Person $\mathcal{B}$ may then engage in the discussion and by trying to persuade $\mathcal{A}$ using either moralized or non-moralized counterarguments. Both theoretical perspectives described above 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 original argument's moral framework are going to be *more* persuasive than non-moralized or morally incongruent counterarguments.

The present paper tests both competing predictions by analyzing online discussions on the Reddit community `/r/ChangeMyView`.

3   The Subreddit "ChangeMyView"

Reddit is an online discussion board that is organized in 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 each discussion thread by stating their personal opinion on a chosen issue. Other users are invited to challenge the original poster's (OP) arguments and try to persuade them otherwise. The OP then responds to the arguments brought forward and explicitly identifies forum entries that changed his or her view by awarding a "Delta" (Δ). The participation rules of the community foster civil exchange of arguments—even for divisive issues—and OPs are encouraged to award Δs genuinely.[1] To date, the Subreddit has more than 500,000 subscribed 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 to correct for spelling errors):

> *CMV: I believe that the gay marriage discussion isn't as important as the media portrays it to be.*

---

[1]The current set of rules for original posts as well as responses can be accessed at https://www.reddit.com/r/changemyview/wiki/rules.

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 Δ 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 Δ:

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 Δ. 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.

In a recent proceeding of the International Conference on the World Wide Web, 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 compield 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 characterics (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.
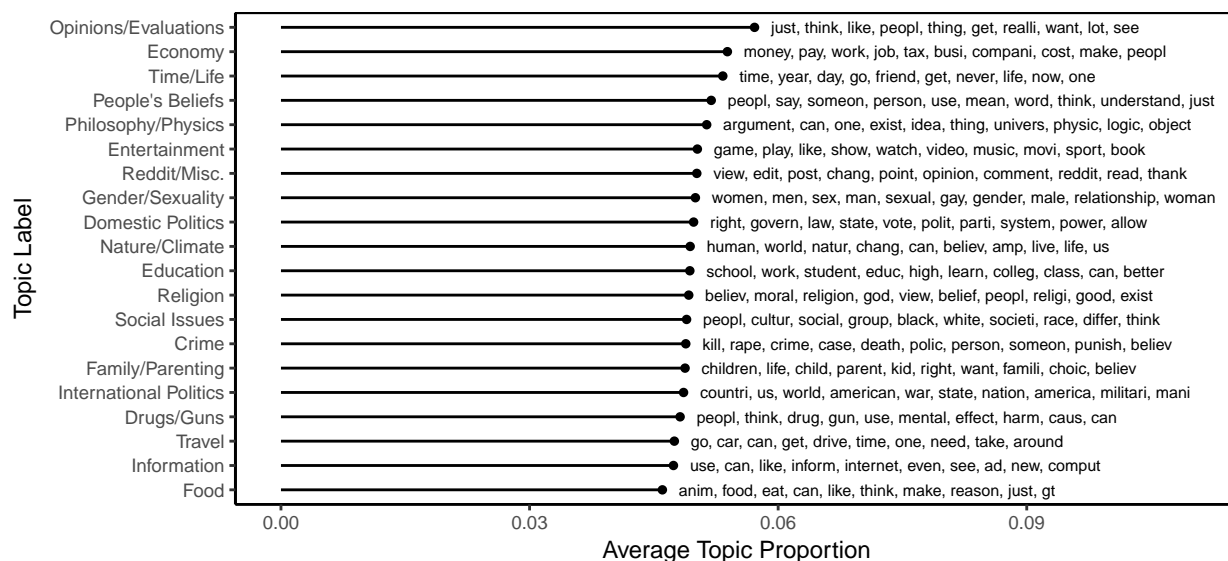

Figure 1: Average topic proportions in original posts on /r/ChangeMyView/ based on a basic LDA model with 20 topics. The plot additionally displays the ten most likely terms associated with each respective topic.

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.

4 Can People on the Internet be Persuaded?

As an initial step, I examine the extent to which OPs are willing to award Δs 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.

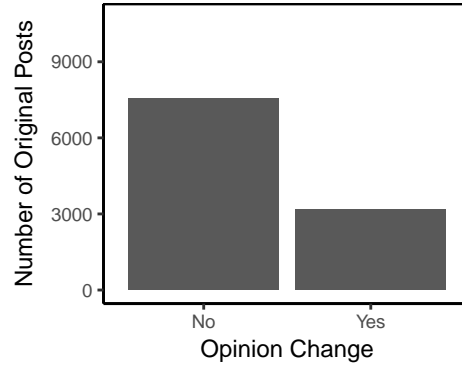In about two thirds of the discussions on CMV between 2013 and 2015, OPs did not award

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

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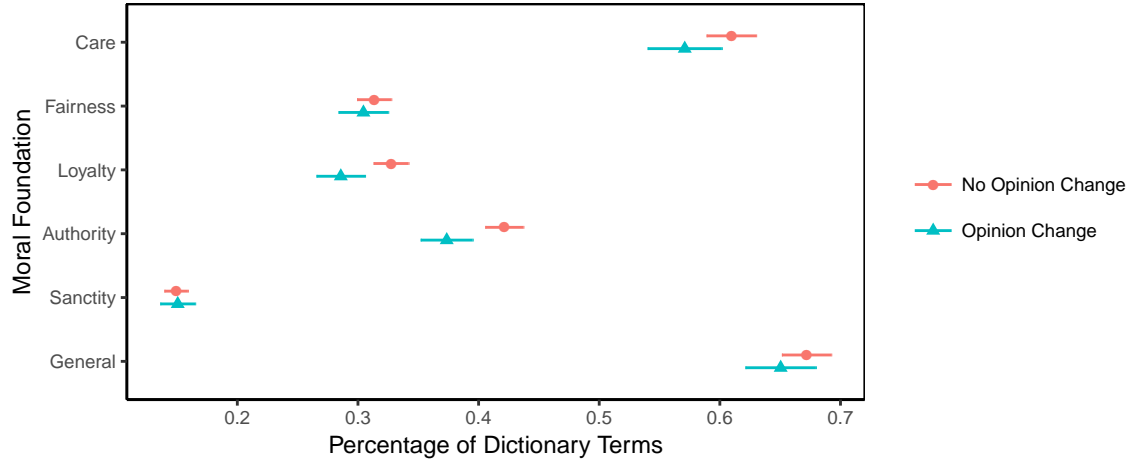
Figure 3: Moral Foundations and Persuadability: Average percentage of dictionary terms relative to the total number of words in each original post starting a discussion, comparing original posts where the author subsequently awarded a Δ (opinion change) or not (including 95% confidence intervals).

after accounting for multiple comparisons using Bonferroni correction). OPs who did not award any Δs 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, 2016).

5 Determinants of Convincing Arguments

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:

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

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 larges proportion of common words (disregarding stopwords).[2] 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 Δ. The response that received a Δ 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.
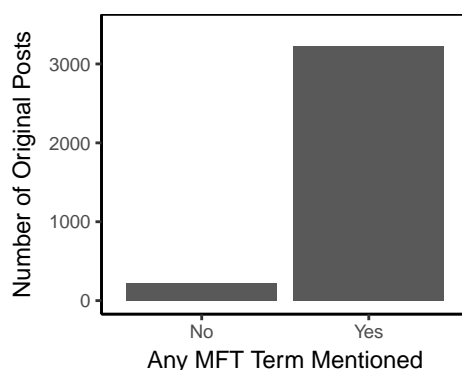

Figure 4: Number of original posts in the paired argument data that included *any* term mentioned in the moral foundation dictionary.

One might worry that this selection on discussions where OPs ultimately awarded a Δ might discard all cases where the initial statement emphasized moral considerations. Luckily, that

---

[2]Elaborate on additional selection criteria.

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 Δ, which might be problematic when examining the effect of moral dictionary terms.
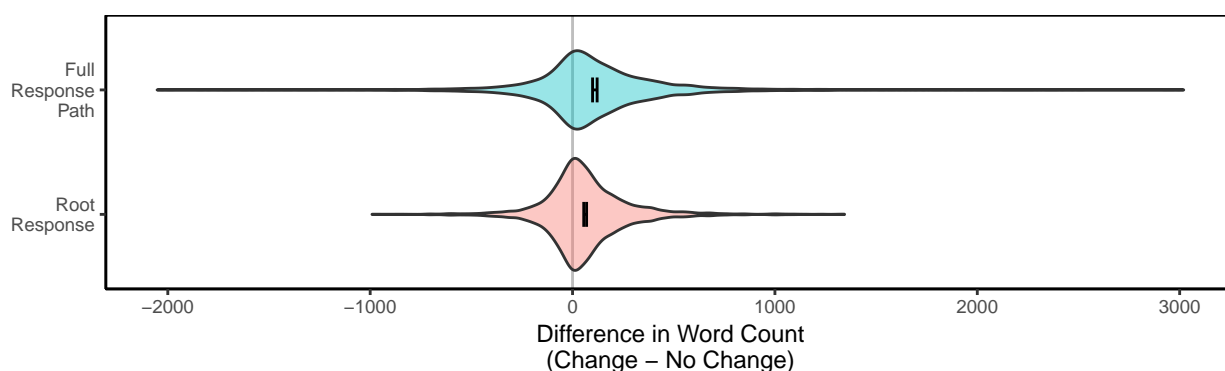

Figure 5: Difference in response lengths between successful and unsuccessful counterarguments. The narrow black bars display the 95% confidence interval of mean differences.

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 igure 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.

5.1 Moral Reasoning and Opinion Change

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 Δ contained a larger percentage of dictionary terms, and vice versa.
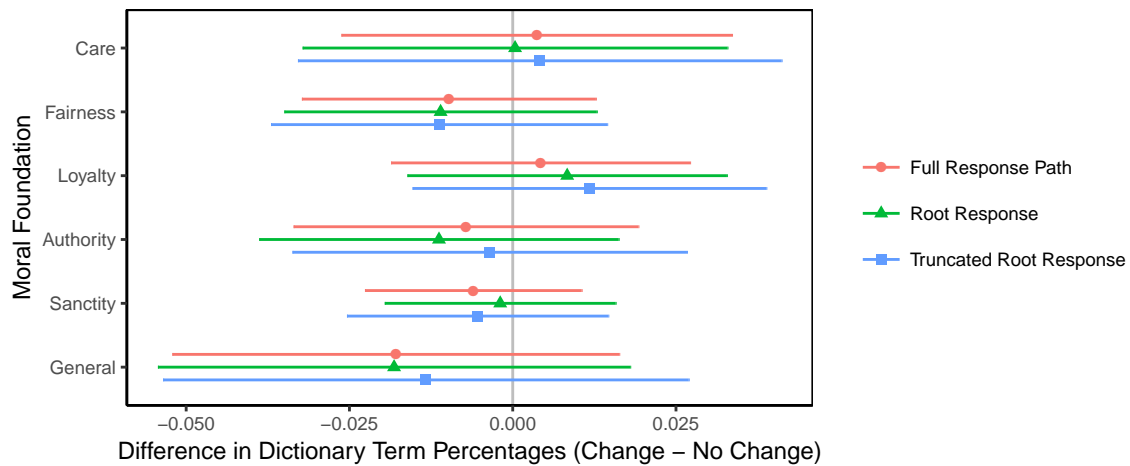

Figure 6: Moral Foundations and Persuasiveness: Average difference of dictionary term percentages relative to the total number of words in each post, comparing counterarguments where the OP subsequently awarded a Δ (opinion change) or not (including 95% confidence intervals).

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.

5.2 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 cosine similarity in dictionary scores between opening statements by OPs and the respective

counterarguments. More formally, moral congruence is written as:

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

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 cosine similarities between successful and unsuccessful counterarguments. Positive values indicate higher congruence in moral appeals among counterarguments that were ultimately awarded a Δ by OPs. The results clearly show that congruence in moral appeals is larger in persuasive than non-persuasive arguments.
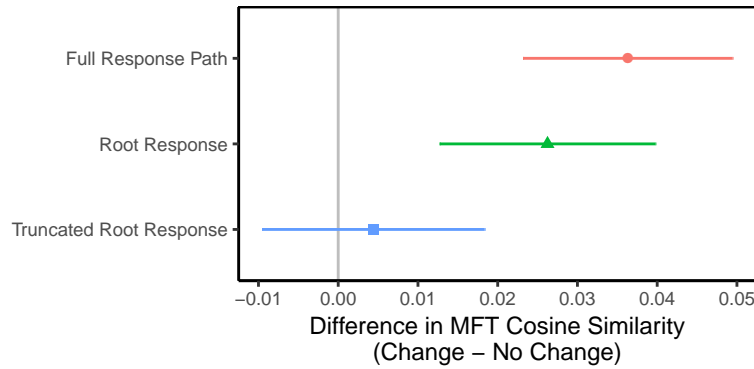

Figure 7: Moral congruence and persuasiveness: Average differences between successful and unsuccessful counterarguments in cosine similarities with the opening statements (including 95% confidence intervals).

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.

6 Conclusion

The preliminary results are consistent with moral foundations theory and contradicts the previous literature on moral convictions. While the general moralization of arguments appeared to had little effects on their persuasiveness, the results show that moral congruence with initial 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.

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

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.

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.

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.

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.

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.

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* forthcoming: 1–11.

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.

Ryan, Timothy J. 2014. "Reconsidering Moral Issues in Politics." *Journal of Politics* 76 (2): 380–397.

Ryan, Timothy J. 2016. "No Compromise: Political Consequences of Moralized Attitudes." *American Journal of Political Science* forthcoming (doi:10.1111/ajps.12248).

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.

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.