

ChangeMyView: Can Moral Appeals Facilitate Compromise?*

Patrick W. Kraft *Stony Brook University*

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

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)

***Version:** March 01, 2018 — PLEASE DO NOT CITE OR REDISTRIBUTE WITHOUT PERMISSION.

(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 related to political attitudes, evaluations, and behavior.

An important implicit assumption that has been made repeatedly 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) argued 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) suggested 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 showed 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 persuasion and the likelihood of compromise: While moral foundations theory contends that agreement should be facilitated if two discussants focus on the same underlying foundations, the moral conviction literature suggests that any type of moral appeal should make it harder to overcome disagreement. The present paper tests both competing predictions by analyzing online discussions on the Reddit community /r/ChangeMyView. The following section discusses the data set in more detail.

Description of Dataset

ChangeMyView (CMV) is a Reddit community where participants begin a discussion by describing their opinion on a chosen issue in an original statement. Other participants are invited to challenge the original poster's (OP) arguments in order to change their view. The OP then responds to the arguments brought forward and explicitly identifies entries that changed his or her view by awarding a "Delta" (Δ). For example, one original post that was added in 2014 (? check date again!) discusses the issue of marriage equality:

"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.

Below is a sample response that lead the original author to award a Δ :

“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.”

The participation rules of the community foster civil exchange of arguments—even for decisive issues—and OP’s are encouraged to award Δ s genuinely. As such, it provides a perfect opportunity to examine the role of moral appeals in persuasion and agreement.

[Tan et al. \(2016\)](#) analyzed the discussion dynamics on ChangeMyView by focusing on language features that predict persuasiveness of discussion contributions as well as the malleability of original posts. The authors extract 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.]

The analyses reported below use the same data and case selection that was published by [Tan et al. \(2016\)](#) as well as the Moral Foundations Dictionary proposed by [Graham, Haidt, and Nosek \(2009\)](#).

Moral Foundations and Persuadability

This figure shows that OPs who use fewer moral words in their opening statements are more likely to provide a Δ in the subsequent discussion. This basic result is consistent with the moral conviction literature.

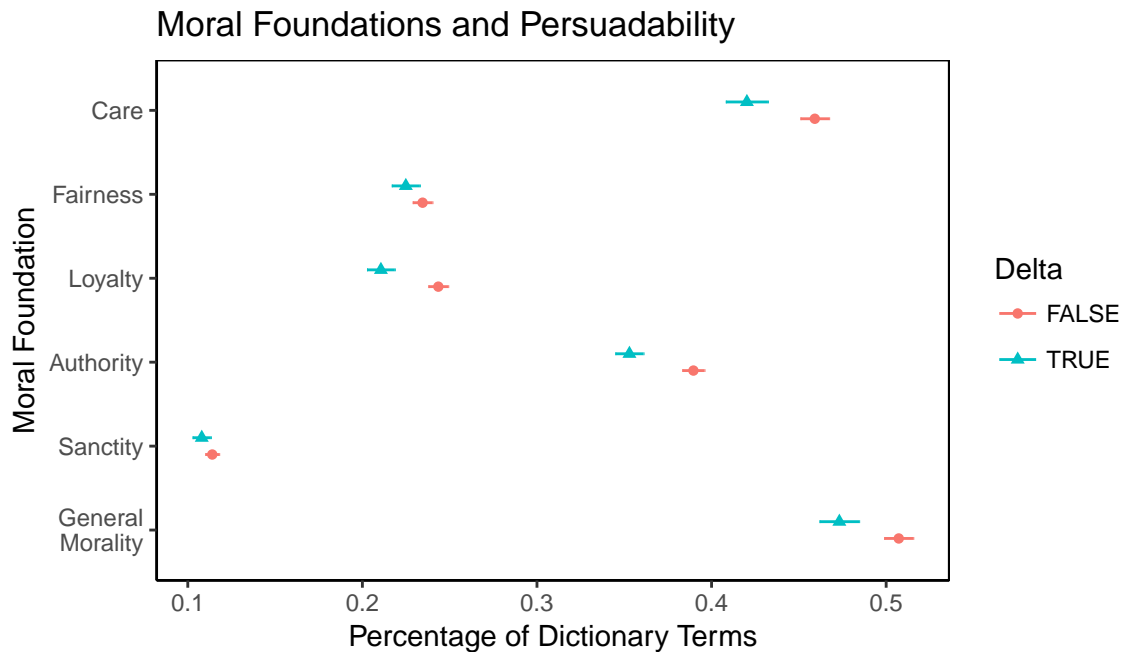


Figure 1: Average percentage of dictionary terms relative to the total number of words in each original post starting a discussion (including 95% confidence intervals). Deltas indicate whether the author was persuaded to change his or her view as a result of the discussion.

Is Moral Consistency Convincing?

Next, I examine the persuasiveness of moral arguments made within discussion contributions as a response to original posts that rely on any of the moral foundations.

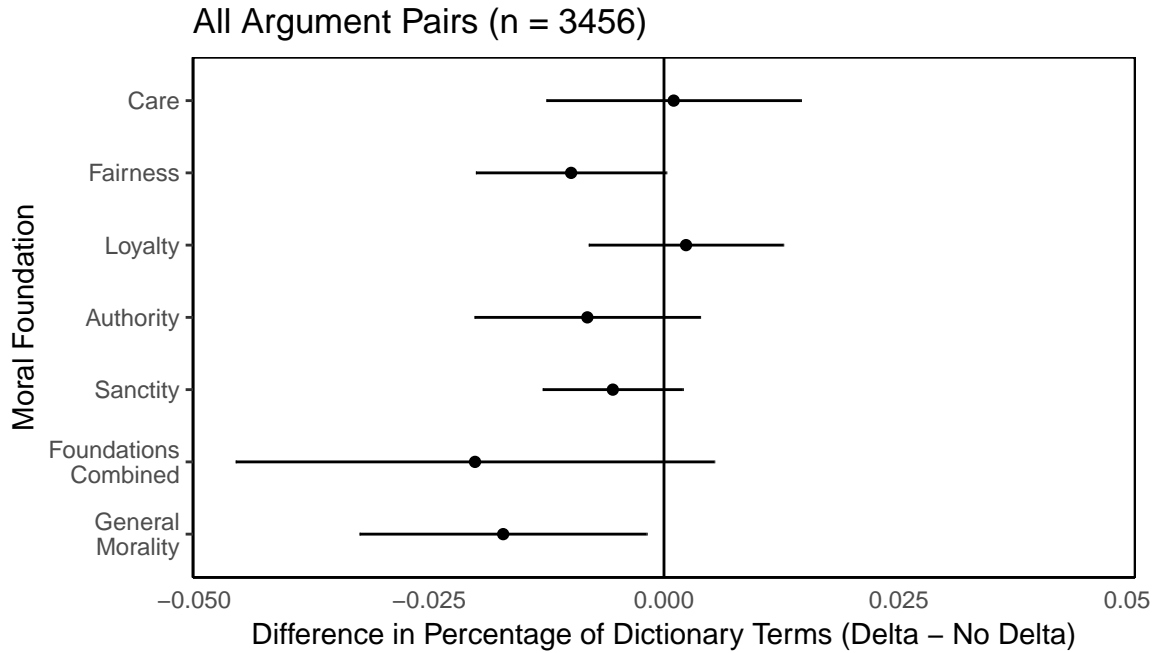


Figure 2: All argument pairs: Average difference in percentage of dictionary terms relative to the total number of words between matched discussion contributions that persuaded the author of the original post vs. not (including 95% confidence intervals). Negative values indicate that arguments were less persuasive (and vice versa).

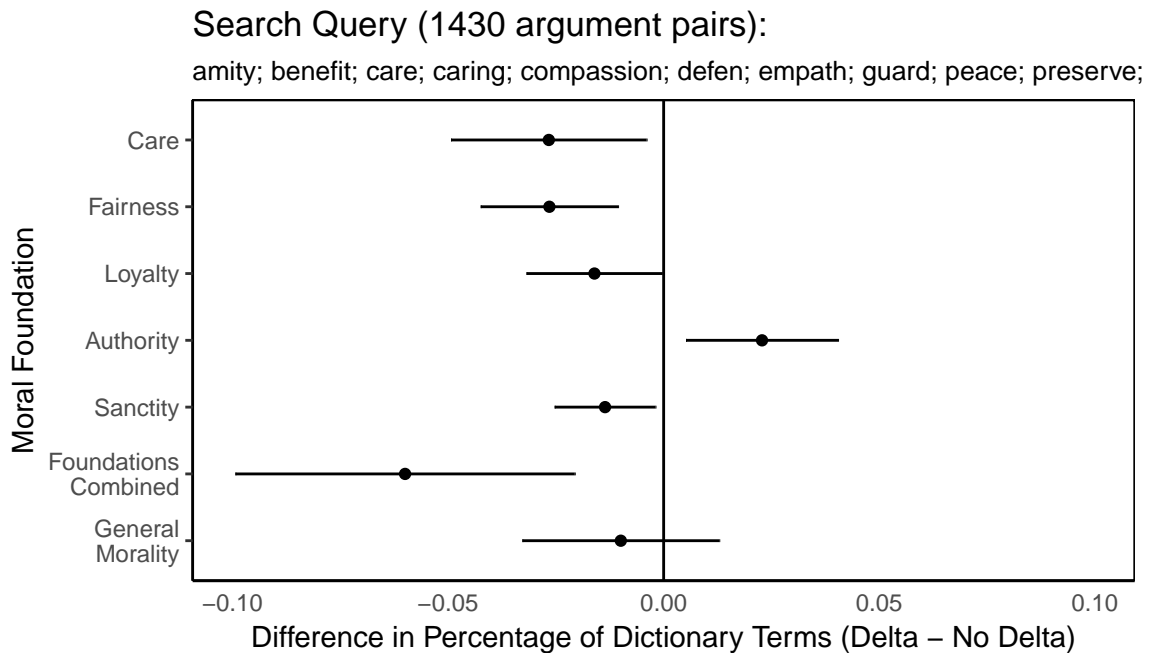


Figure 3: Care (virtue): Average difference in percentage of dictionary terms relative to the total number of words between matched discussion contributions that persuaded the author of the original post vs. not (including 95% confidence intervals). Negative values indicate that arguments were less persuasive (and vice versa).

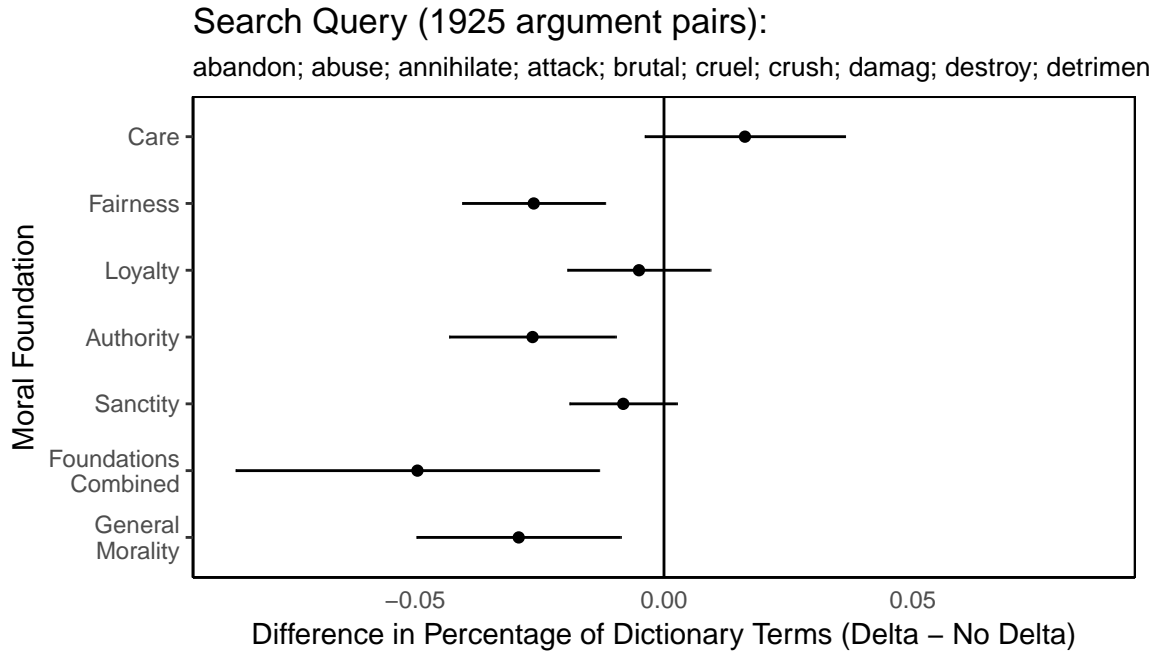


Figure 4: Care (vice): Average difference in percentage of dictionary terms relative to the total number of words between matched discussion contributions that persuaded the author of the original post vs. not (including 95% confidence intervals). Negative values indicate that arguments were less persuasive (and vice versa).

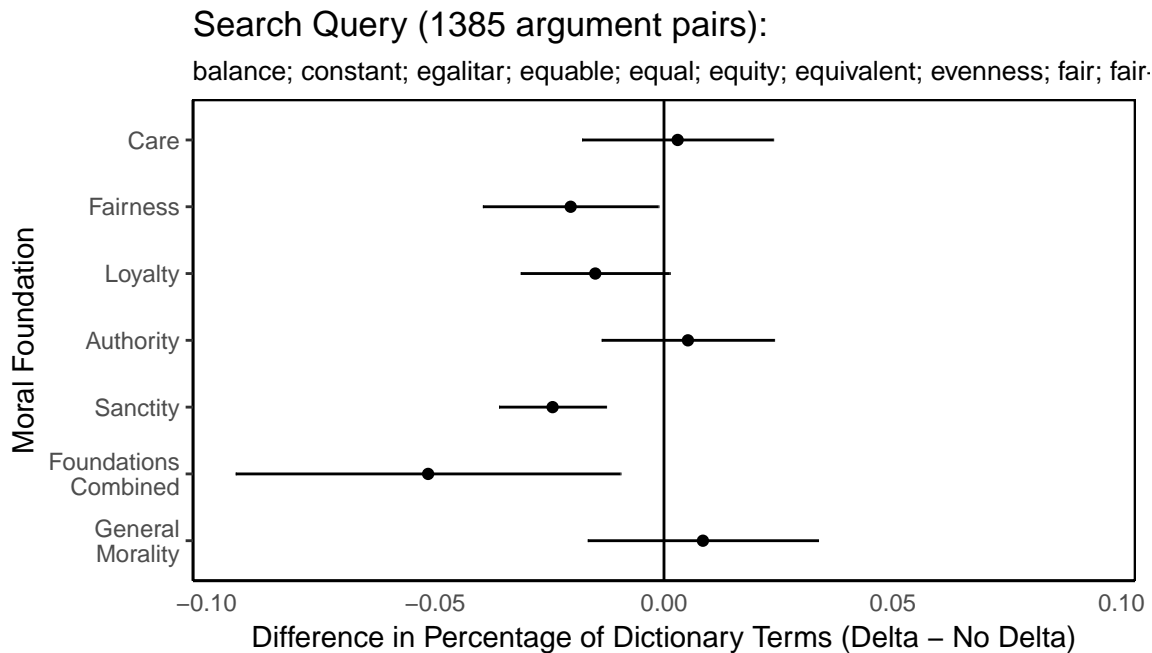


Figure 5: Fairness (virtue): Average difference in percentage of dictionary terms relative to the total number of words between matched discussion contributions that persuaded the author of the original post vs. not (including 95% confidence intervals). Negative values indicate that arguments were less persuasive (and vice versa).

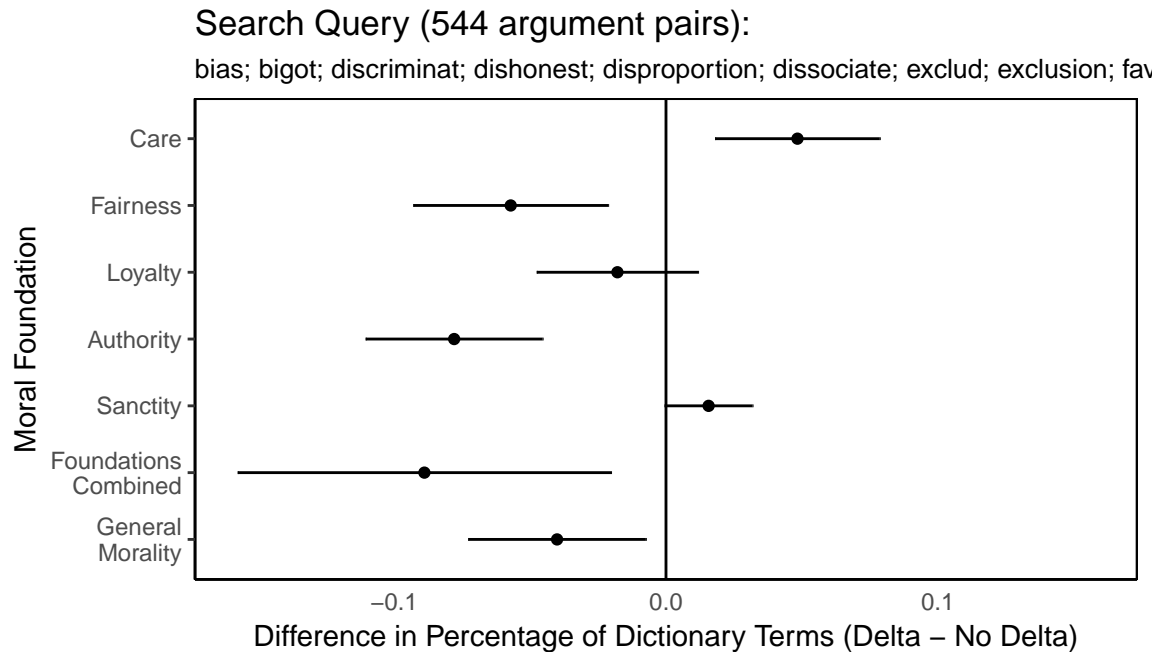


Figure 6: Fairness (vice): Average difference in percentage of dictionary terms relative to the total number of words between matched discussion contributions that persuaded the author of the original post vs. not (including 95% confidence intervals). Negative values indicate that arguments were less persuasive (and vice versa).

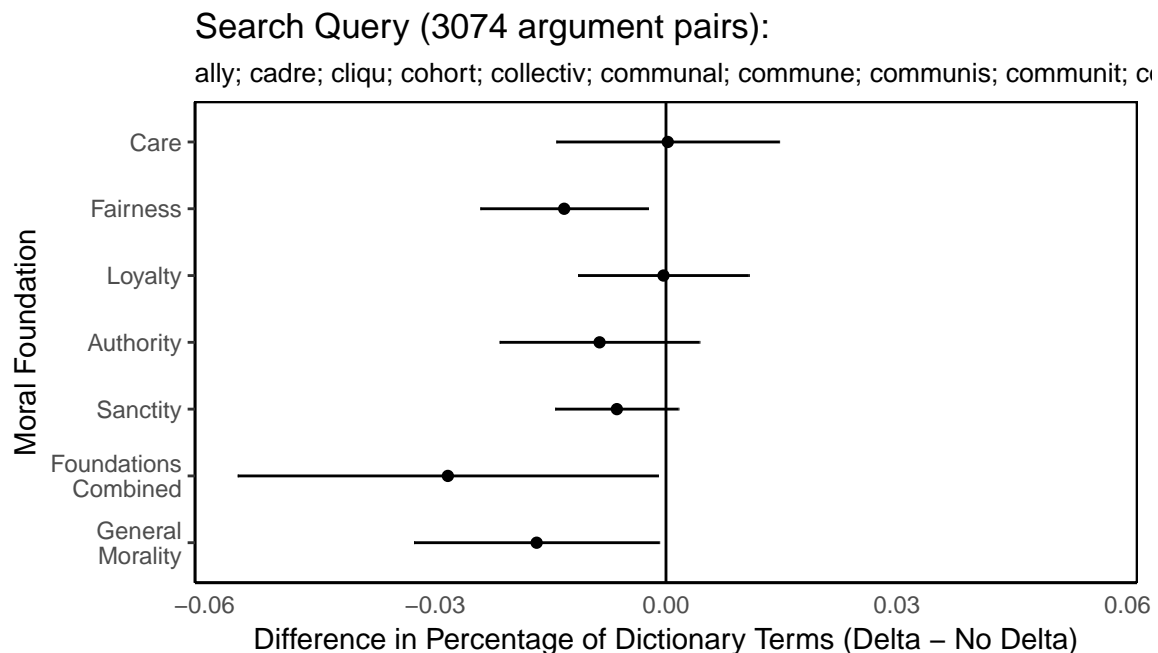


Figure 7: Loyalty (virtue): Average difference in percentage of dictionary terms relative to the total number of words between matched discussion contributions that persuaded the author of the original post vs. not (including 95% confidence intervals). Negative values indicate that arguments were less persuasive (and vice versa).

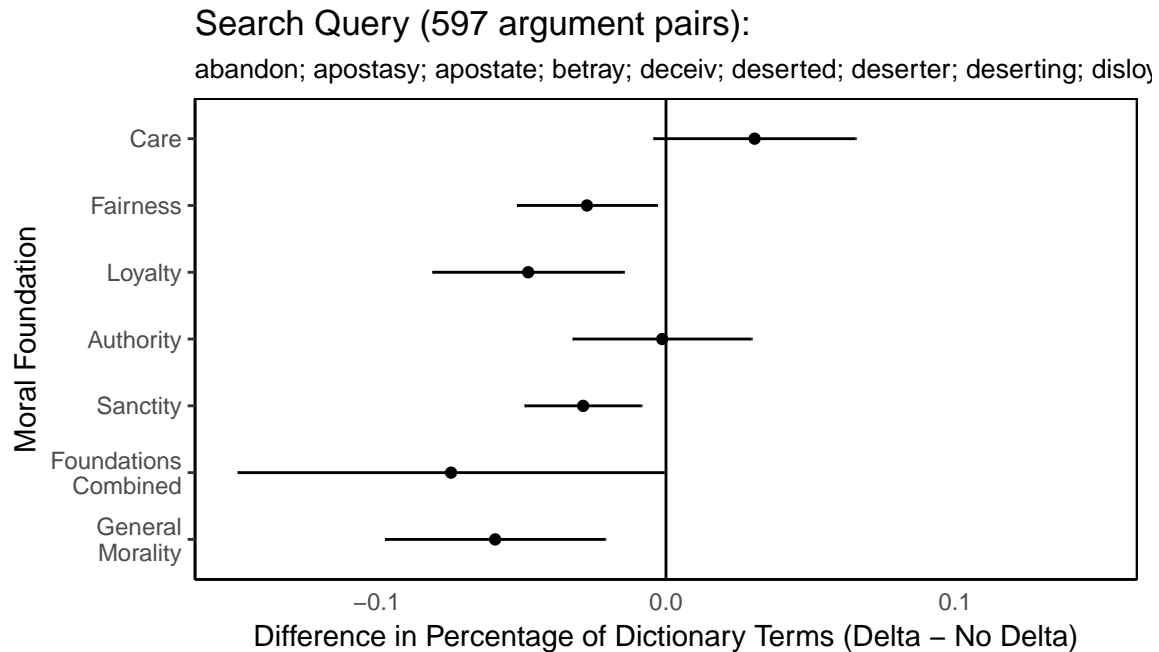


Figure 8: Loyalty (vice): Average difference in percentage of dictionary terms relative to the total number of words between matched discussion contributions that persuaded the author of the original post vs. not (including 95% confidence intervals). Negative values indicate that arguments were less persuasive (and vice versa).

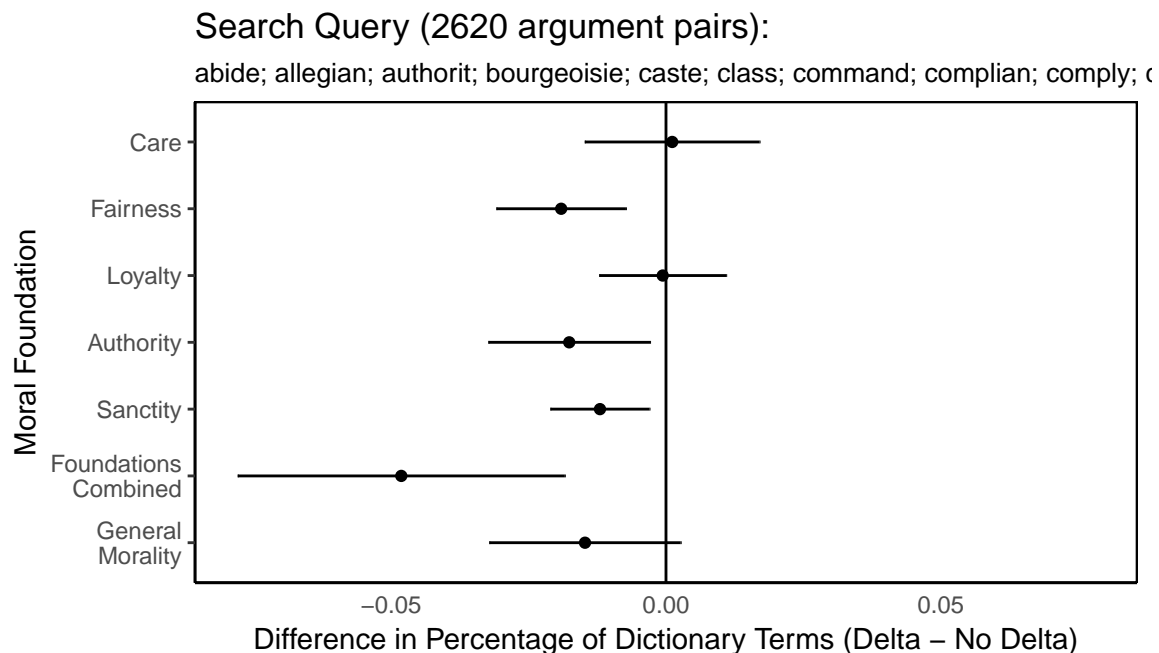


Figure 9: Authority (virtue): Average difference in percentage of dictionary terms relative to the total number of words between matched discussion contributions that persuaded the author of the original post vs. not (including 95% confidence intervals). Negative values indicate that arguments were less persuasive (and vice versa).

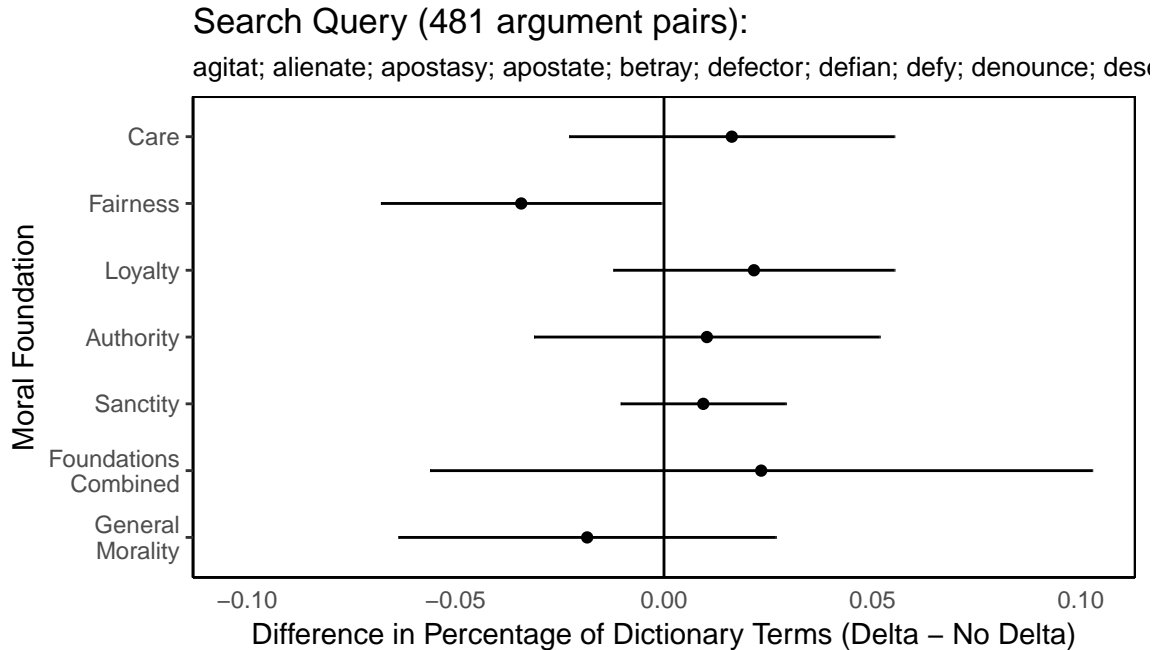


Figure 10: Authority (vice): Average difference in percentage of dictionary terms relative to the total number of words between matched discussion contributions that persuaded the author of the original post vs. not (including 95% confidence intervals). Negative values indicate that arguments were less persuasive (and vice versa).

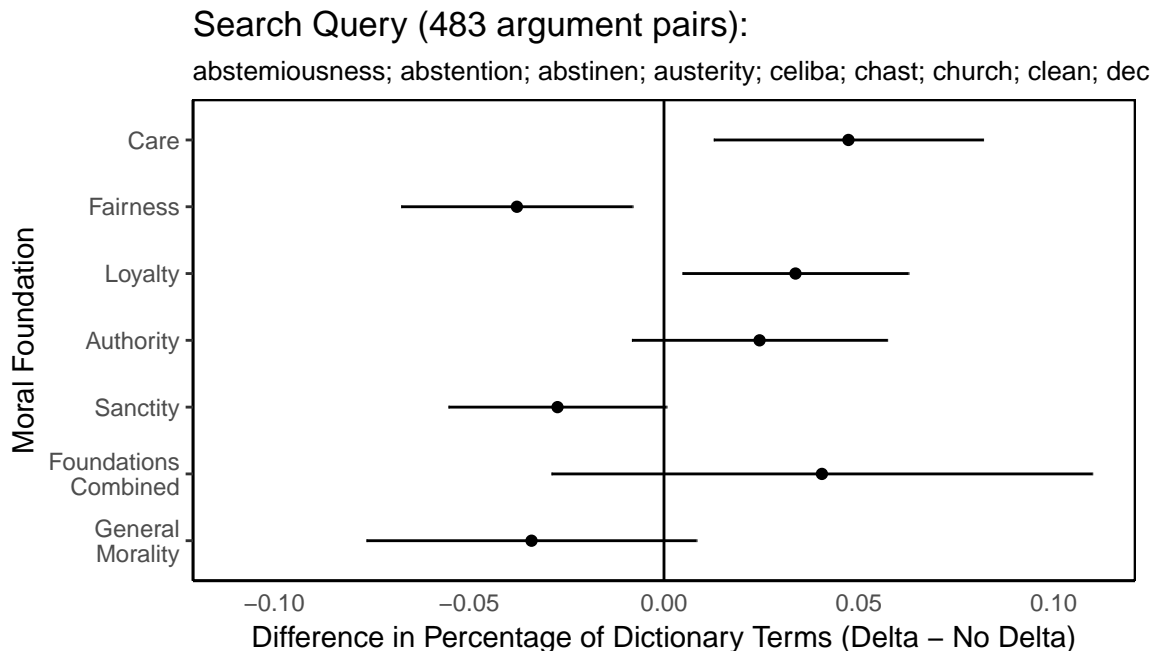


Figure 11: Sanctity (virtue): Average difference in percentage of dictionary terms relative to the total number of words between matched discussion contributions that persuaded the author of the original post vs. not (including 95% confidence intervals). Negative values indicate that arguments were less persuasive (and vice versa).

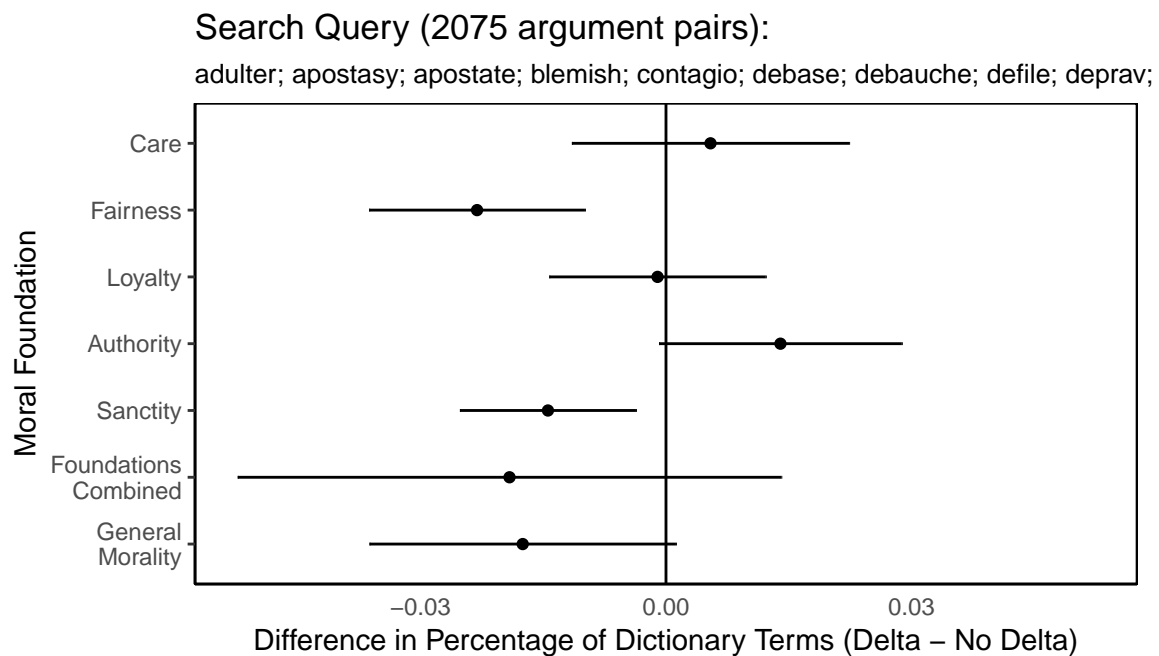


Figure 12: Sanctity (vice): Average difference in percentage of dictionary terms relative to the total number of words between matched discussion contributions that persuaded the author of the original post vs. not (including 95% confidence intervals). Negative values indicate that arguments were less persuasive (and vice versa).

Conclusion

The preliminary results are consistent with the literature on moral conviction and clearly contradict the implicit assumptions made in the Moral Foundations literature. Overall, arguments that revolve around the same moral foundations are less likely to be persuasive. In most cases, moral appeals of any type appear counterproductive in facilitating compromise and changing people's minds.

TO DOs, future directions, etc.

- examine specific topics (e.g., climate change etc.)
- clean original posts (links etc)
- adjust confidence intervals to correct for multiple comparisons
- improve code documentation, add comments in internal functions
- check MFT scores in argument pairs (as well as OP entry)

References

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