Is it fair? Textual effects on the salience of moral foundations

Eyal Sagi (esagi@stfrancis.edu)
University of St. Francis
500 Wilcox St. Joliet, IL, 60435

Abstract

Many of the important decisions we make have moral implications. Moral Foundations Theory (Haidt & Joseph, 2004) identifies 5 distinct styles of moral reasoning that may be applied to such decisions. This paper explores how reading text that emphasizes one of these styles might affect our reasoning. After participants read a series of tweets that emphasized the Fairness/Cheating foundation they exhibited an increased reliance on this style compared to when they read tweets emphasizing the Care/Harm foundation. This affected participants' answers to a questionnaire designed to measure the perceived importance of the different foundations, as well as in their rating of the foundations evident in other tweets. Interestingly, this effect was short lived and was not observed for the Care/Harm foundation. These results suggest that exposure to the moral reasoning of others might temporarily influence what moral arguments we are likely to accept and employ.

Keywords: Framing; Moral Foundation Theory; Moral Cognition; Priming; Text

Introduction

Many of the important decisions we make have moral implications. But what factors might affect these decisions? In this paper, I examine the effect that encountering moral arguments might have on subsequent reasoning about moral issues. More specifically, I will argue that moral reasoning is subject to priming effects, where being confronted with a particular style of moral reasoning will result in increased salience for that style of reasoning.

Moral Foundations Theory

While psychological research on morality encompasses a wide range of theoretical approaches (e.g., Gray, Young, & Waytz, 2012; Malle, Guglielmo, & Monroe, 2014; Rai & Fiske, 2011; Young & Saxe, 2011), in this paper I am interested in comparing different styles of moral reasoning and will therefore focus on Moral Foundations Theory (Graham et al., 2013; Haidt & Joseph, 2004). Moral Foundations Theory identifies five different types of moral intuitions or concerns: Harm, Fairness, Loyalty, Authority, and Purity. Each of these moral concerns accounts for a different style of reasoning about moral dilemmas.

For instance, consider a person who believes that climate change is a problem because it endangers the lives of people and animals. This person is primarily concerned with the *harm* that climate change could cause to living beings. In contrast, another person might argue that climate change is problem because of its complexity and global reach, making it the obligation of nations to adhere to guidelines set by international treaties. That person is using a type of argument that emerges from reasoning about *authority*. Critically, when analyzing any argument, it is important to remember that such moral concerns are not exclusive, and that a single argument can exhibit traits from several different concerns.

Priming Moral Reasoning

Research based on Moral Foundations Theory has demonstrated that sensitivity to the different moral concerns varies across cultures (Graham, Haidt, & Nosek, 2009), as well as based on ideological beliefs (Graham et al., 2009; Koleva, Graham, Iyer, Ditto, & Haidt, 2012). Much of this research implicitly assumed that these styles of reasoning are stable and related to personality traits and beliefs. However, many stable traits in psychology provide a baseline for behavior that is affected by contextual and situational factors, such as priming.

The study presented here is designed to test whether such factors can also affect the salience of individual foundations. Specifically, I hypothesize that exposure to moral ideas and beliefs will result in temporary changes to the salience of the foundations that are at the core of these ideas.

For example, if an individual is presented with a text that relies on reasoning based on fairness, this individual might then become sensitized to the Fairness/Cheating foundation and be more likely to consider it as an important aspect of other, more ambiguous lines of reasoning. Likewise, reading a text about an individual that is harmed by a callous individual is likely to predispose the reader to identify harm as a more relevant consideration in subsequent texts that they otherwise would have.

To test this prediction, participants will be presented with a series of tweets that endorse either the Care/Harm foundation or the Fairness/Cheating foundation. Following this presentation, they will be asked to complete tasks that are designed to measure their sensitivity to these concerns. If moral reasoning is subject to priming effects, it is expected that participants who were presented with tweets endorsing the Fairness/Cheating foundation would find issues of fairness to be more relevant and important. In contrast, participants who read tweets that highlight Care/Harm should show heightened concern for that foundation.

Method

Participants

Thirty-six native English speakers from the University of St. Francis participated in the study in exchange for course credit.

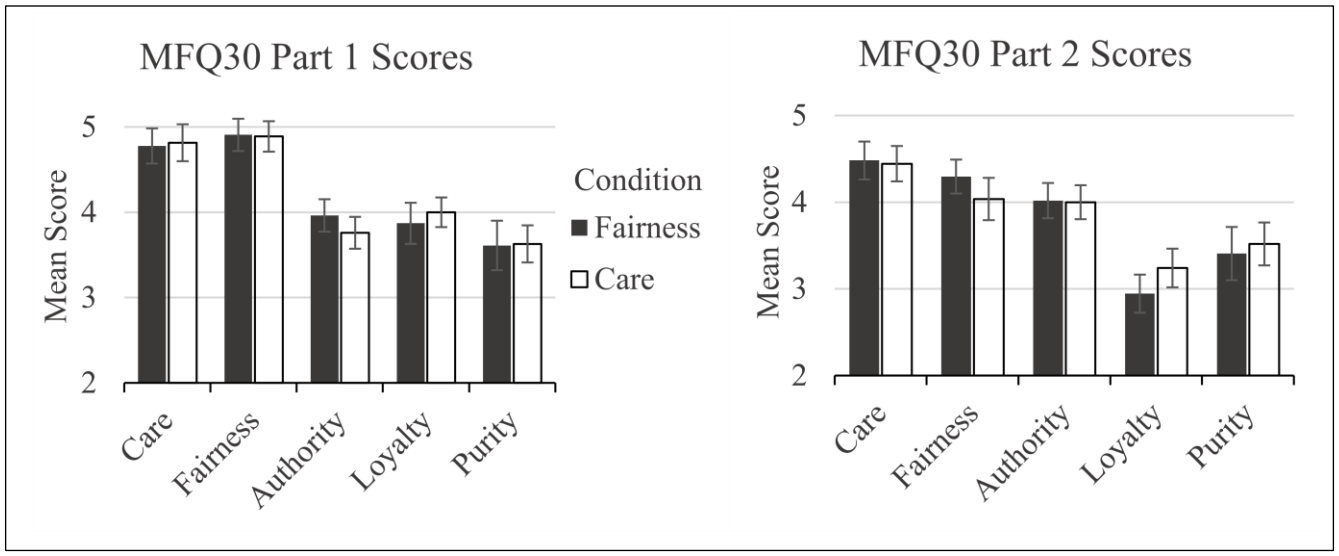

Figure 1 - Mean scores on both parts of the MFQ30, by priming condition. The prime is administered after part 1 and before part 2. Error bars represent standard of error of the mean.

Materials

Moral Foundations Questionnaire

One of the frequently used tools for assessing an individual's level of concern for each of the 5 foundations is the 30 item Moral Foundations Questionnaire (MFQ30; Graham et al., 2011).

This questionnaire is composed of 2 parts: The first part asks participants to rate the relevance (on a 6-point scale) of various considerations to whether an act is right or wrong (e.g. "Whether someone suffered emotionally"). The second part asks participants to rate their agreement (on a 6-point scale) to various statements (e.g. "Chastity is an important and valuable virtue"). Each part is comprised of 16 items, 3 items corresponding to each of the foundations and 1 catch item.

It is important to note that while the two parts are measuring the same underlying concepts, they are using different approaches and therefore the scores on one part of the MFQ are not directly comparable to scores on the other. Nevertheless, a high score on a particular foundation in the first part can be taken as an indication of high concern for that foundation, and is therefore predictive of the score on the second part.

In this study, I used the first part of the MFQ30 to establish a baseline profile of the participants and the second part (administered after the care or fairness prime) to test for a priming effect.

Tweets

In addition to the Moral Foundations Questionnaire, this study made use of several sets of tweets. These tweets were chosen from a corpus of over 700,000 tweets about the U.S. Federal Shutdown of 2013 (cf. Dehghani et al., 2016; Sagi & Dehghani, 2014b)[1]. Tweets were selected based on ratings of moral language computed statistically based on the Moral Foundations Dictionary (Graham et al., 2009) following the method described in Sagi and Dehghani (2014a).

The first set of primes, used as the prime in the Care condition, were uniformly high on the foundation of Care/Harm and low on the other 4 foundations. Likewise, a second set of primes served as the prime in the Fairness/Cheating condition. These primes were high on fairness and low on the other 4 foundations. Each of these lists comprised of 14 tweets, 7 tweets from liberal users and 7 from conservatives (see Appendix A).

In addition, a list of 25 tweets was selected such that each foundation was represented by 5 tweets. As before, for a foundation to be so represented, the tweet had to rate high on that foundation and low on all other foundations. This list of tweets was used for the rating task.

Procedure

Participants first completed the first half of the 30 item Moral Foundations Questionnaire (MFQ30; Graham et al., 2011). Next, they rated their agreement, on a scale of 1 to 6, to a series of 14 tweets that emphasized either the Fairness/Cheating foundation (Fairness condition) or the Care/Harm foundation (Care condition). After rating the primes, they completed the second half of the MFQ30.

Finally, each of the 5 moral foundations was described to the participants using the text from the website *moralfoundations.org* and they were asked to rate, on a scale of 1 to 6, the relevance of each of the foundations to 25 tweets. Of the 25 tweets, 5 were primarily associated with each of the foundations. The tweets were presented in a

[1] This corpus was used because it was pre-analyzed and the ratings were successfully used in previous studies.

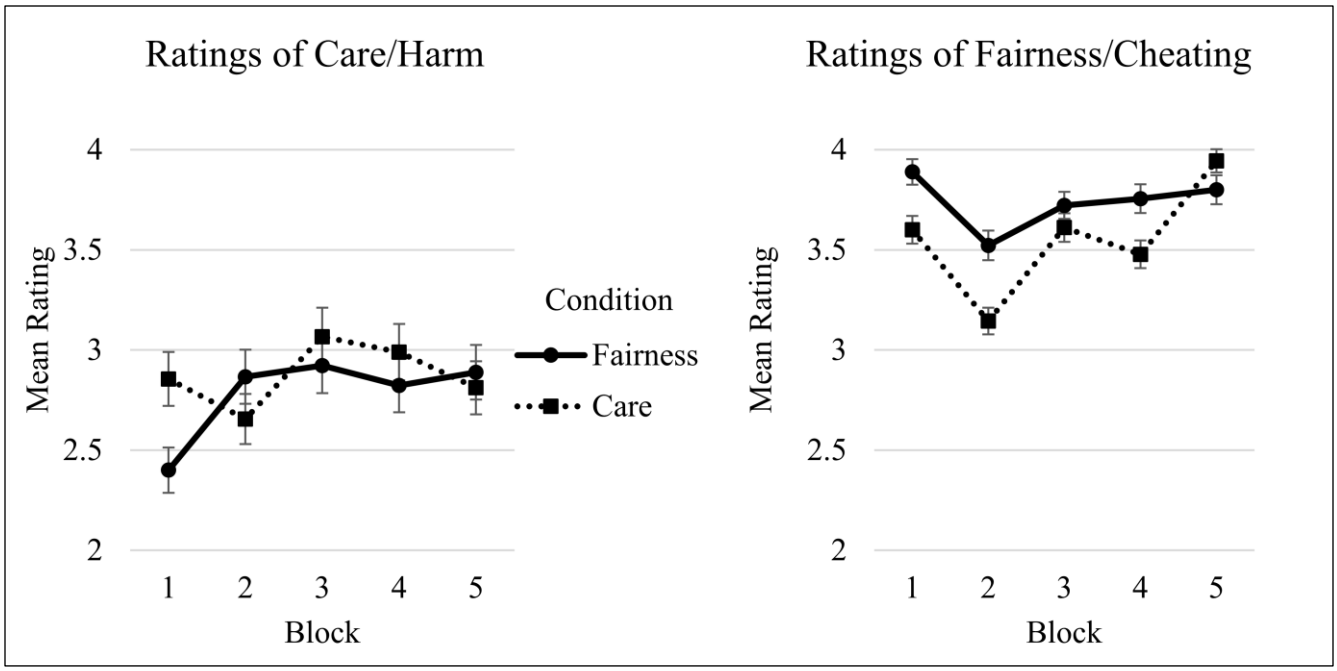

Figure 2 - Mean ratings of tweets on the foundations of Care/Harm and Fairness/Cheating by experimental condition and order. Each block represents 5 tweets, in order of presentation. Error bars represent standard of error of the mean.

random order and its reverse, counterbalanced across participants. The ordering of the tweets (i.e., whether presented in the original random order or the reversed order) did not affect any of the analyses.

Results

Moral Foundations Questionnaire

Figure 1 presents the mean scores on both parts of the questionnaire. Since the prime is only presented after participants complete the first part, no differences are predicted in it. Agreement with the primes did not significantly differ based on condition (Care: $M = 3.69$, SD = 0.68; Fairness: $M = 3.61$, $SD = 0.63$; $t(34) = -0.36$, *n.s.*)

Participants' responses to the second part of the MFQ30 were analyzed using a separate general linear model for each of the foundations, with the foundation score in the first part of the MFQ30 and the prime condition (Fairness vs. Care) as independent variables[2]. Participants' scores on the second part of the questionnaire were correlated to their scores on the first part, at least marginally (after correcting for multiple tests), for all but the Loyalty/Betrayal foundation ($F(1, 32) = 1.87$, $p = .18$; $F(1, 32) > 6$, $p < .05$, $r^2 > .19$ for all other foundations). Only the Fairness foundation showed the predicted interaction between the score on the first part of the MFQ30 and condition ($F(1, 32) = 10.00$, $p < .01$, $r^2 = .34$; Fairness condition: $M = 4.30$, SD = 0.78; Care condition: $M = 4.04$, SD = 0.98; $F < 1$ for all other foundations).

Moreover, the scores on the *fairness* questions of the second part of the MFQ30 of participants in the Care condition correlated with their fairness score on the first part ($r(16) = .73$, $p < .001$) while those of participants who read tweets evoking fairness did not ($r(16) = -0.11$, *n.s.*). This suggests that following the fairness primes participants concern for fairness was uniformly high – the prime essentially set all participants to the same level of concern on fairness. In contrast, similar correlations on the harm scores of the MFQ30 did not differ significantly (care condition: $r(16) = .42$, $p = .08$; fairness condition: $r(16) = .63$, $p < .01$). These results suggest that the fairness prime successfully increased the salience of the Fairness/Cheating foundation, the care prime did not increase the salience of the Care/Harm foundation.

Ratings of Tweets

To simplify the analysis of the ratings and avoid repeated tests, the analysis of the 25 rated tweets used a single model that contrasted the ratings of harm and fairness (although a similar, post-hoc, model using all 5 foundations yielded qualitatively similar results). This model included the participant and the tweet as random variables, and the condition as well as the foundation being rated as independent variables. To test for the possibility that this effect diminishes over time, the 25 tweets were divided into 5 blocks of 5 tweets based on order of presentation and this variable was included in the model (model $r^2 = .37$). As

[2] Since the hypothesized effect is due to mere exposure to the text, participants' agreement with the primes was not predicted to affect the results and it is therefore omitted from the analysis. Importantly, models that include this variable show no effect of agreement and are otherwise unchanged.

predicted, participants rated tweets as higher in fairness if they were previously exposed to tweets that exhibited fairness-based reasoning and vice versa ($F(1, 1734) = 5.46$, $p < .05$). However, this effect quickly diminished as is evident by its interaction with the order of presentation ($F(1, 1734) = 3.88$, $p < .05$; see Figure 2).

Discussion

The results of the present study demonstrate that reading texts that evoke principles of moral reasoning can affect judgments and decisions made later. The effects observed in this paper are therefore best considered to be a type of priming effects. Since priming effects are, for the most part, short lived, the rapid decay of the effect in the second part of the study is also easily explained. However, it is likely that, because the second part of the study overtly asked participant to consider all five styles of moral reasoning, it accelerated this decay and that in a more natural setting the effect might last longer.

Perhaps more interesting is the fact that while reading tweets involving fairness and cheating resulted in a priming effect, reading tweets that favored the foundation of Care/Harm did not. One possible explanation is that while the federal shutdown readily appealed to the foundation of Fairness/Cheating, its appeal to Care/Harm is less direct and evident. This is reflected in the tweets – although Care/Harm was a dominant foundation in the corpus for liberals, considerations of fairness dominated the overall debate (cf. Sagi & Dehghani, 2014b). It is possible that rather than simply evoking a moral foundation, a consistent and/or clear moral position might be required for a text to affect the moral reasoning of its reader.

More generally, there are numerous studies that demonstrate how the use of language can affect reasoning, both in the lab (e.g., Tversky & Kahneman, 1981), and outside of it (e.g., Goodwin, 1994). Moreover, it is possible to use language to measure and trace the history of such frames (Sagi, Diermeier, & Kaufmann, 2013).

In a similar vein, there is evidence that situational factors affect an individual's moral reasoning. The bystander effect, where individuals are less likely to render assistance when there are many other bystanders than when there are few, is a prominent example of such an effect (Darley & Latane, 1968).

This study combines these two well-known effects and demonstrates that this type of framing can provide a context in which moral reasoning takes place. More interestingly, it is possible that repeated exposure to particular styles of reasoning might have a cumulative effect and eventually lead to the salience of the relevant foundation being permanently increased (or, perhaps, decreased, depending on the circumstances of exposure). This type of effect might be at the root of the development of moral beliefs and might provide insight into how and why such beliefs change.

Moreover, even temporary effects might have important implications. For example, the language used to draft jury instructions might influence the verdict one way if it highlights fairness and another if it highlights care. Likewise, during negotiations, it is possible that a particular choice of language and reasoning by one side can serve to focus the negotiation in a particular direction, influencing all parties towards emphasizing the importance of a specific concern.

References

Darley, J. M., & Latane, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, *8*(4, Pt.1), 377–383. https://doi.org/10.1037/h0025589

Dehghani, M., Johnson, K., Hoover, J., Sagi, E., Garten, J., Parmar, N. J., … Graham, J. (2016). Purity homophily in social networks. *Journal of Experimental Psychology: General*, *145*(3), 366–375. https://doi.org/10.1037/xge0000139

Goodwin, C. (1994). Professional vision. *American Anthropologist*, *96*(3), 606–633.

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S., & Ditto, P. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, *47*, 55–130. https://doi.org/10.1177/0963721412456842

Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, *96*(5), 1029–1046. https://doi.org/10.1037/a0015141

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the Moral Domain. *Journal of Personality and Social Psychology*, *101*(2), 366–385. https://doi.org/10.1037/a0021847

Gray, K., Young, L., & Waytz, A. (2012). Mind Perception Is the Essence of Morality. *Psychological Inquiry*, *23*(2), 101–124. https://doi.org/10.1080/1047840X.2012.651387

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*, *133*(4), 55–66.

Koleva, S. P., Graham, J., Iyer, R., Ditto, P. H., & Haidt, J. (2012). Tracing the threads: How five moral concerns (especially Purity) help explain culture war attitudes. *Journal of Research in Personality*, *46*(2), 184–194.

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A Theory of Blame. *Psychological Inquiry*, *25*(2), 147–186. https://doi.org/10.1080/1047840X.2014.877340

Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: Moral motives for unity, hierarchy, equality, and proportionality. *Psychological Review*, *118*(1), 57–75. https://doi.org/10.1037/a0021867

Sagi, E., & Dehghani, M. (2014a). Measuring Moral Rhetoric in Text. *Social Science Computer Review*, *32*(2), 132–144. https://doi.org/10.1177/0894439313506837

Sagi, E., & Dehghani, M. (2014b). Moral Rhetoric in Twitter: A Case Study of the US Federal Shutdown of 2013. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 1347–1352).

Sagi, E., Diermeier, D., & Kaufmann, S. (2013). Identifying Issue Frames in Text. *PloS One*, *8*(7), e69185.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. https://doi.org/10.1126/science.7455683

Young, L., & Saxe, R. (2011). When ignorance is no excuse: Different roles for intent across moral domains. *Cognition*, *120*(2), 202–214. https://doi.org/10.1016/j.cognition.2011.04.005

Appendix A – Primes

Care/Harm Primes

Dr. Seuss's #Congress Who Stole #SNAP: Kids seniors face health risks due to #GovernmentShutdown. #PublicPolicy

New #Obama Doctrine: Protect oil, allies, the homeland from terrorists weapons of mass destruction.

#governmentshutdown Day 9. Private charity pays Military death benefits instead of #Pentagon. What do you think?

As #soldiers we're told #WWII, #Vietnam, #Iraq, were all to protect #Democracy, yet the latest attack comes from The #TeaParty #Shutdown

#Shutdown: #Obama Keeps Military #Golf_Courses #Open, #Closes Military #Grocery_Stores |

Mother of fallen #soldier denied death benefits: #Criminal to treat our #soldiers this way #congress. via @todayshow

If #obama can treat our military and vets like he does imagine how he's going to be with civilians and our healthcare. Disgust

#Obama is trying his hardest to create pain - Military death benefits denied to families of fallen troops

I don't care about the shutdown...PAY the families of our fallen heroes!!! #shutdown #governmentshutdown #Military #veterans #Obama

I wish #Congress cared as much about war vets benefits as they do about the war vets memorial. #hypocrite #pander #teaparty

Sickening that the families of our fallen heroes denied benefits by shutdown. Time to stop the madness.#shutdown

Outrageous not paying death benefits to families of our fallen servicemen! This SOB #Obama looking for a civil war to become dictator !

The D-Day memorial in Normandy, France has been closed, upsetting tourists and veterans. via @WSJ #shutdown

Refugees Waiting Overseas Are in Limbo as U.S. Shutdown Continues #refugees #shutdown #resettlement #newcomers #USA

Fairness/Cheating Primes

Liberal #Congress members claim that the law must apply equally to all...well, except them. #Obamacare #Dems #GOP

Libs scream #obamacare = law of the land. Weird cuz theyre VERY WILLING 2 ignore immigration borders, ya kno another LAW OF THE LAND

A bunch of liberals looked really stupid tonight, talking about #obamacare. They're still ignorant of the law. #tcot #election2014 #pathetic

Fighting Republican hysteria with calm analysis on the ACA. #p2 #toppage #dems #liberals #progressives #healthcare

Hey #GOP! #OBAMACARE website overloaded huh? Looks like Americans want an alternative you elephant sized asses!

I think it's hilarious T-Party called ACA #Obamacare as a negative slur. The more popular it gets, the bigger my SMILE gets POTUS's too!

Hey #GOP look up the 14th amendment! If u love the Constitution Founding Fathers so much, then ADHERE to the law of the land. #JustVote

Liberal #Congress members claim that the law must apply equally to all...well, except them. #Obamacare

Equal under the law; all laws enforced equally - its pretty simple for everyone to understand except Obama #TeaParty #tcot #tngop #gop #ccot

Y did the unions get exempt from #Obamacare I thought it was the law of the land doesn't it apply to everyone like every other law #tcot

Funny how libs like @tamaraholder are all about #obamacare being the law but other social issues like upholding the sanctity of marriage..

LIberals progressives say that #obamacare is the law of the land, but they ignore illegals breaking the law of the land!

also calling progressives 'liberals' (not saying that someone IS liberal, but calling them 'liberals') is #GOP branding.

Smart Libs know Repubs hate that #Obama wins. He beat them twice in elections, SCOTUS upheld #ACA. It just kills em. 2BAD!