The Moral Foundations in New York Times Editorials

Ashley Dainas
Indiana University, USA
ardainas@umail.iu.edu

Vipul Munot
Indiana University, USA
vipmunot@gmail.com

Satoshi Tsutsui
Indiana University, USA
stsutsui@indiana.edu

Abstract

Finding moral words and distinguishing the kinds of moral reasoning used in a text is an underdeveloped area of research with many potential applications. The present study attempts to apply Jonathan Haidt's Moral Foundations Theory, a psychological theory of human morality that proposes the idea that there are five basic modules of morality, to a computer science context. We describe our creation of a pilot corpus of NYT articles from January 2015, calculate corpus statistics for the data, and analyze these preliminary findings. We find that the Care/Harm and Authority/Subversion foundations were used most in both comments and articles. We also found that when we normalized the moral foundation word use, commenters and articles had very similar mean scores. Finally, we discuss potential areas for future work in this area.

1 Introduction

The advent of the internet, Web 2.0 and social media have allowed for the creation of a massive amount of data. The data created are many and varied, but the most obvious data is linguistic. The internet does allow for multimodal communication through images, video and even audio, but the primary mode of discourse is text based. Researchers from many different fields, including Information Science, Cognitive Linguistics, Psychology and Computer Science, have taken advantage of this boon and developed methods such as sentiment analysis (Pang & Lee, 2008) and emotion analysis (Aman and Szpakowicz, 2007) in order to better understand how people convey opinions and emotions.

These methods use theories from the social sciences, such as Ekman's (1992) theory of basic emotions, to analyze online data and then train algorithms on that analysis. This practice has been incredibly effective at enabling the analysis of data from social media and other online sources. The area of Moral Psychology has unfortunately not yet benefited from this treatment. This paper seeks to argue that Jonathan Haidt's (2012) Moral Foundations Theory can be used for a "moral analysis" of online data.

In order to begin illustrating the validity of this argument, we conducted a pilot study based on a small corpus collected from the New York Times Online Editorial pages from January 2015. We analyzed the corpus using the Moral Foundations Dictionary (Graham, Haid, & Nosek, 2009) and calculated the normalized corpus statistics for the five foundations and general morality for both the articles and the comments. We have also started some preliminary work in analyzing the stability of moral foundation use in highly committed New York Times Online commenters.

2 Related Work

Jonathan Haidt's Moral Foundations Theory is a theory of human morality that posits the idea that "morality is organized in advance of experience" (Haidt & Joseph, 2004; Haidt, 2012). This means that humans are born with a preinstalled intuitive ethics. The word "intuitive" is drawing on a psychological theory that divides the mind into "two systems" (Kahneman, 2011). System 1 is the fast, intuitive system that relies heavily on built-in processing, so-called "gut-instincts" and heuristics. System 2 is the slow, deliberate system that requires conscious effort and is used when we think explicitly in words or try to reason out the answer to a problem. According to this theory, System 1 makes intuitive snap decisions and judgements in about a second or two, leaving System 2 to make up reasons to support System 1's conclusions.

The systems can be seen at work in things like job interviews where it has been shown that interviewers decide within seconds of shaking the interviewees hand whether or not they'll be hired even though the interviewer is sure that the one hour interview was vital. System 1 made the snap judgement and System 2 spent an hour or more coming up with good reasons why the already made decision was a good one (Kahneman, 2011). It also determines personal preferences about matters such as media consumption and interpersonal relationships. Haidt's theory suggests that morality works the same way. We are born with an intuitive, System 1 based ethics, which means we are wired to feel flashes of approval or disapproval to specific social patterns. Haidt argues that this intuitive system is the foundation for all moral systems in the world. However, just because all people have the same intuitive ethics at birth, it does not mean that all people use the ethics the same way. Morality is socially constructed in that children are taught by their parents, teachers and culture which moral foundations to obey and why. Thus, System 1 is honed and System 2 learns which moral feelings to listen to and how to justify them (Haidt & Joseph, 2008).

There are currently six moral foundations that have been confirmed, through meta-analyses of years of moral psychological studies as well as various experiments and survey studies (Haidt & Joseph, 2004; Haidt & Graham, 2007; Haidt, 2012; Graham, Koleva, Iyer, Haidt, Motyl, Wojcik, & Ditto, 2013). The foundations are Care/Harm, Fairness/Cheating, Loyalty/Betrayal, Authority/Subversion, Purity/Degradation, and Liberty/Oppression. Below is a brief description of each foundation:

1. Care/Harm deals with concerns about the suffering of others and the desire to prevent harm from happening. This foundation probably originates from parental instincts to care for crying offspring. Those instincts have been expanded to include people outside of the family, the animals and even the environment.

2. Fairness/Cheating deals with issues of proportionality and justice.

3. Loyalty/Betrayal is related to tribal structures, in-group/outgroup mentalities and the obligations of group membership.

4. Authority/Subversion concerns the social order, hierarchical relationships, obedience and respect. This foundation is all about obeying tradition and legitimate authority.

5. Purity/Degradation is related to the emotion of disgust and is triggered by physical and spiritual contagions. This foundation can be seen when conservatives express concerns over interracial marriage and when liberals worry about how they will be harmed by consuming genetically modified food.

6. Liberty/Oppression is the most recently accepted moral foundation (Iyer, Koleva, Graham, Ditto & Haidt, 2012). Its proposed mechanism is that when an alpha male mistreats the weaker members of the group, these weaker members can band together and defeat the oppressor. This foundation concerns the hatred of tyranny and the love of freedom. It was discovered because it is the only foundation the Libertarians care about.

The sixth foundation, Liberty/Oppression, will be omitted from the analyses in this study because it was not a part of the MFD (Graham, Haid, & Nosek, 2009). The MFD contains 295 words and word stems related to the first five foundations as well as some general moral words. Each of the five foundations is split into “Virtues”, which are positive terms associated with a foundation, and “Vices”, which are negative words associated with a foundation.

Being able to understand, access and even predict a person’s moral preferences based on their writing has many potential applications in politics, psychology, public health, and social justice among many other areas. Haidt’s research (Graham, Haidt & Nosek, 2009) has found that liberals and conservatives differ in which of the moral foundations they endorse different moral foundations. Liberals consistently relied almost entirely on the Care/Harm and Fairness/Cheating foundations (this was before the Liberty/Oppression was separated from Fairness/Cheating). Conservatives on the other hand, endorsed all five foundations equally. Haidt proposes that awareness of the differences in moral foundations use between liberals and conservatives can encourage understanding and easier communication between the two groups rather than the usual instances of both sides speaking past each other with incompatible moral arguments. If this is true, a user’s moral foundations use might predict what kind of politics or religious views they may have. This knowledge can be used to craft arguments aimed specifically at convincing either liberals or conservatives. Of course, Haidt et al do take a very strong stance on this argument. Frimer et al (2013) have shown that while the moral codes of liberals and conservatives do differ systematically, their similarities outweigh their differences at least in one area. Frimer found that liberals and conservatives both use Care/Harm, Fairness/Cheating, and Purity/Degradation when making moral judgments about influential people. Only further study will show just how similar and different liberals and conservatives are across the foundations in various circumstances.

Another potential application for Moral Foundations Theory is that it can be used to assess moral rhetoric and see how people respond to and are convinced by moral rhetoric. We may be able to trace how moral foundations are used throughout an evolving discussion and see what, if anything, changes a person's moral foundation use. Using moral foundations on big data or even small data could open a great many doors for bettering our understanding of the moral mind.

Little work has been done on using Moral Foundations Theory to analyze big data, but there have been three recent studies that consider the problem. Vaisey & Miles (2014) did not actually use the theory themselves, but they did suggest that Moral Foundations Theory could be imported into Sociology and used to analyze interview, archival material and big data. Sagi & Dehghani (2014) attempted to measure the moral rhetoric related to specific topics, in this case, documents written before and after the '93 and '01 attacks on the World Trade Center, the conflict over the ''Ground-Zero Mosque'', and the abortion debate in the US Senate. They used a computational text analysis technique called latent semantic analysis which computes semantic similarity between concepts in the text being analyzed and MFD keywords. Their technique can be used for the rapid analysis of moral rhetoric in online texts. The third and final big data study, sentiment analysis study that has been done using Moral Foundations Theory (Dehghani, Sagae, Sachdeva & Gratch, 2014) examines liberal and conservative rhetoric around "Ground Zero Mosque" conflict using three different text-processing algorithms. The authors found that conservatives and liberals were most distinguished by the words they used related to negative portrayal of outgroups. Conservatives and liberals were most similar in terms of the Virtue or positive aspects of Greater similarity in Virtue domains of Vice for Care/Harm, Fairness/Cheating and Authority/Subversion.

The moral analyses described above illustrate just how promising Moral Foundation Theory is. There is a lot that can be done in taking the theory beyond its moral psychological roots and using it in Information Science and Linguistics research.

3 Data

Due to the fact that Moral Foundations Theory has not yet been applied to social media there are no available datasets to work from. The only publically available document is the official Moral Foundations Dictionary (Graham, Haid, & Nosek, 2009). The dictionary contains 295 words and word stems related to the first five moral foundations as well as some general moral words. Each of the five foundations is split into "Virtues", which are positive terms associated with a foundation, and "Vices", which are negative words associated with a foundation. We originally want to take a machine learning approach to classifying the moral foundations. Unfortunately, there was no readily available training data that identifies the political leanings of users. Thus, we decided to forgo a machine learning approach in this exploratory study. Instead we utilize corpus statistical methods to begin to get an idea of where the Moral Foundations Theory may best be applied in the future.

We initially wanted to use data pulled from social media such as Tumblr, Reddit and Twitter. Unfortunately, the APIs of Tumblr and Reddit were not conducive to collecting posts with high moral language content. We decided that Twitter was not suitable for this pilot study because we felt the 140-character limit would not allow for enough moral content for the posts to be worth studying. We did however observe that the posts that did get the most commentary on social media tended to lead back to news sources and major blogs. These posts tended to have dozens, if not hundreds of comments depending on the topic. We considered several news sites such as the Washington Post and the New York Times (NYT). Ultimately we selected NYT opinion and editorial pieces from January 2015 for this study due to the ease of the data's availability.

The NYT has an API (http://developer.nytimes.com/docs) that we used to retrieve the data from a list of article URLs for a particular period. The comments for an article can also be retrieved this way. The biggest challenge was that it was not possible to fetch the full text of the articles by API. To overcome this issue we implemented an HTML scraper to pull the text for us. We used Python to implement each step of the data collection process in order to create our pilot corpus.

The corpus consists of 602 articles and 80,358 total comments. The data was collected from the NYT online articles from January 2015.

4 Method

In order to analyze our NYT corpus for moral foundation use, we first needed to find the best method of stemming our data so that it could be matched against the words in the MFD. We conducted a pilot study on one article and a portion of its comments in order to see how Moral Foundation Theory can be applied to our data and discover what methods to use in analyzing the corpus. We assumed that a NYT article was morally controversial if it had a high number of comments.

One annotator went through the text of the article and first 60 comments by hand and noted down the number of MFD words from each of the eleven categories using the “Find” tool in Microsoft Excel. The annotator also recorded the word count for each article and comment.

The hand annotated results were compared against the results of using five different stemmers on the same data. The stemmers tested were the Porter stemmer, Stanford stemmer, Lancaster stemmer, Snowball stemmer, and Wordnet stemmer. The Stanford NLP stemmer utilized python wrapper since it was developed in Java and the other four came from the NLTK library. Of the five stemmers, the Porter stemmer results matched the hand annotated results the best and we judged it to be the best stemmer for our needs.

We applied the Porter stemmer to all of the articles and comments and then counted up each of the moral foundations words. We also recorded the user IDs, user locations and post IDs that signify when comments are responses to other comments.

We utilized a K-means algorithm using R and its inbuilt library for k-means algorithm in an attempt to identify whether or not moral foundation use formed clusters. Then, we used Python to calculate basic corpus statistics for the data, finding the mean, median, mode, maximum and minimum use of each category across both the articles and the comments. . We used Python’s Matplotlib and seaborn packages to plot the statistics and to plot the individual foundations in order to observe the variations in the foundations.

Based on personal observations, we hypothesized that that there would be a small number of users contributing a large amount of comments and a large number of users only contributing a small amount of comments. In order to see if this was true, we extracted all of the users from the corpus. We counted how many comments each user made. Then, we extracted each users’ moral foundations for all their comments across all articles for January 2015. We calculated the basic statistics of the two biggest power users of January 2015.

5 Results

Table 1 records the mean, median and maximum values for each moral foundation as well as the word count across all 602 articles. The mode and median values were omitted because they were simply 0. On average, Care/Harm was the most used moral foundation (1.71) closely followed by Authority/Subversion (1.41). General moral words were used the most on average (1.73). The articles were on average 636 words long. The highest maximum number of general moral words in an article was 40 words. The maximum number of Care/Harm words that appeared in an article was 28 and the maximum number of Authority/Subversion words was 25.

	Care	Fair	Loyal	Auth	Purity	Moral	WC
Mean	1.71	0.25	0.99	1.41	0.20	1.73	636.3
Med.	1	0	0.5	0.5	0	1	616.5
Max	28	14	17	25	15	40	3794

Table 1: Moral foundation statistics for 602 articles from January 2015.

Table 2 records the mean, median and maximum values for each moral foundation as well as the word count across all 80,358 comments. Again, the mode and median values were omitted because they were simply 0. On average, Care/Harm was the most used moral foundation (0.24) closely followed by Authority/Subversion (0.18). General moral words were used the most on average (0.34). The comments had a mean word count of 83 words, though the maximum word count recorded for a comment was 293. The highest maximum number of general moral words in a comment was 14 words. The maximum number of Care/Harm words that appeared in a comment was 20, and the maximum number of Fairness/Cheating words was also 20.

	Care	Fair	Loyal	Auth	Purity	Moral	WC
Mean	0.24	0.04	0.09	0.18	0.03	0.34	82.73
Med.	0	0	0	0	0	0	64
Max	20	20	13	11	11	14	293

Table 2: Moral foundation statistics for 80,358 comments from January 2015

Knowing the average number of moral words used in an article or comment is valuable. However, the information can be misleading. If we compare a comment that has 100 words and uses 10 Care/Harm words against a 200 word comment that had the same number of Care/Harm words, and we only consider the count, then the two comments look the same. However, when we consider the proportion of Care/Harm words to the total number of words, it is clear that the first comment has a much higher proportion of Care/Harm words over the second, since a proportion of 0.1 is greater than 0.05. Thus, we normalized the mean and mode across all comments and all articles.

Table 3 records the normalized mean and maximum values for each moral foundation across all 602 articles. On average, Care/Harm was the most used moral foundation (0.00296) closely followed by Authority/Subversion (0.00214). The normalized mean use of general moral words fell just behind Care/Harm (0.00265). The foundation with the highest normalized maximum rate of moral words was Care/Harm (0.071).

	Mean	Maximum
Care/Harm	0.00296	0.071
Fairness/Cheating	0.00035	0.011
Loyalty/Betrayal	0.00161	0.067
Authority/Subversion	0.00214	0.031
Purity/Degradation	0.00029	0.035
General Morality	0.00265	0.039

Table 3: Normalized moral foundations use in 602 articles from January 2015.

Table 4 records the normalized mean and maximum values for each moral foundation across all 80,358 comments. On average, Care/Harm was the most used moral foundation (0.00271) closely followed by Authority/Subversion (0.00202). The normalized mean use of general moral words was incredibly high (0.336835). The foundation with the highest normalized maximum rate of moral words was Authority/Subversion (0.4).

The K-means clustering experiment failed to distinguish any significant patterns in the data.

	Mean	Maximum
Care/Harm	0.00271	0.3333
Fairness/Cheating	0.00051	0.25
Loyalty/Betrayal	0.00102	0.3333
Authority/Subversion	0.00202	0.4
Purity/Degradation	0.00041	0.3333
General Morality	0.336835	0.5

Table 4: Normalized moral foundations use in 80,358 comments from January 2015.

The user who made the most comments on the NYT articles in January 2015 was user id 37674938. This user made 236 comments total comments. The top fifty users made 4089 number of comments in total which accounts for 5.09% of the total number of comments. The trend of the number of comments made by users decreases sharply and then gradually from the maximum in a reverse J-shaped frequency distribution (Figure 1).

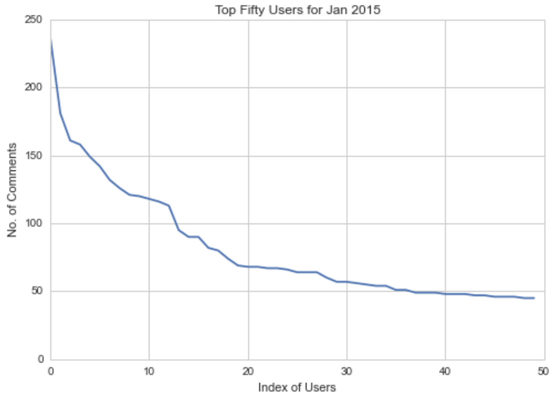

Figure 1: The number of comments by the top fifty users for January 2015. See Appendix I for a larger version of this chart.

The highest number of comments on an article was 1691. The fifty most commented articles had 36097 number of comments altogether which accounts for 44.92% of the comments made in January 2015. The trend of the number of comments per article decreases sharply and then gradually from the maximum in a reverse J-shaped frequency distribution (Figure 2).

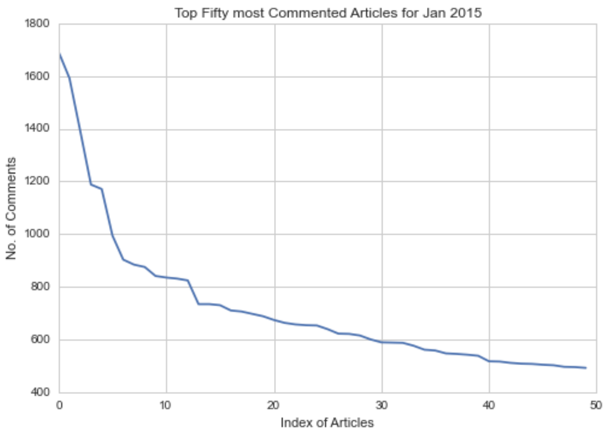

Figure 2: The articles from January 2015 with the highest number of comments. See Appendix I for a larger version of this chart.

6 Discussion and Conclusions

As mentioned in the previous section, Care/Harm was, on average, the most commonly used moral foundation. This held true in both the comments and the articles, as well as when the foundation counts were normalized against total word count. The second most used foundation, on average, was Authority/Subversion. This also held true across comments and articles, normalized data and not. This finding is interesting, because the New York Times is widely considered to be an incredibly liberal publication. Thus, according to Haidt's research (Graham, Haidt & Nosek, 2009), Care/Harm and Fairness/Cheating should be the most commonly used foundations in this context.

Instead, Fairness/Cheating was consistently used only slightly more than the least used moral foundation of Purity/Degradation. In fact, the order of highest use across articles and comments was: Care/Harm, Authority/Subversion, Loyalty/Betrayal, Fairness/Cheating and Purity/Degredation. The dataset used in this exploratory study is small, but this consistent deviation from what morality use Haidt predicts we should find is worthy of further investigation. Future work should investigate whether or not this pattern holds true across multiple years of NYT data. Another avenue of study would be to see what the normalized moral word use is across many different types of media platforms including traditional news sources as well as social media platforms.

The non-normalized statistics show that the NYT articles and comment sections have a wide moral range. An opinion piece might have no moral words at all. Or it may include a huge amount of moral words from one or more of the foundation categories. Despite this wide variation, the data shows that on average, the normalized mean moral foundation use matches very closely across articles and comments. For example, the normalized Care/Harm word use was 0.00296 in the articles and 0.00271 in the comments. The normalized Authority/Subversion word use was 0.00214 in the articles and 0.00202 in the comments. This close relationship between moral word use in the articles and comments held true for all five foundations, though not for general moral word use. This finding suggests that there is some relationship between the articles and the commenters. However, it is unclear if commenters choose to comment on articles that match their moral preferences or if commenters adjust their moral foundation use to match the moral foundations used in the discussion they are joining.

In the future, we would like to follow up on this initial finding by creating six dimensional moral vectors for each commenter and article. We could then run comparisons to see if commenters only comment on articles that match the commenter's moral vector. We could also find out whether or not a commenter's moral vector changes across all their comments, or if their moral foundation use is fairly consistent. This would also allow us to compare the moral foundation use between so-called "power users" and more casual NYT commenters.

This pilot study is very limited in scope. We only looked at one month's worth of articles and comments from one website. Due to time constraints and the limitations of the data, we were only able to complete a basic corpus analysis of the collected data. In the future, we would like to overcome these limitations and find data that has been labeled with users' political affiliations so that we can compare moral foundation use across liberals and conservatives. This would be one way to bring machine learning into the study of moral foundations in social media. One way to get this data would be to pick a highly divisive public issue such as the Syrian Refugee Crisis or the upcoming presidential election and use the declared loyalties of users on various platforms to label the data. This would allow us to train an algorithm to assign political affiliations based on moral foundations.

Moral Foundations Theory proposes that all of human morality has its basis in a series of moral foundations that every human is born with. Although the theory is well supported by survey and experimental methods, there has been very limited work done to apply Moral Foundations Theory in a computer science context. In this paper we have proposed that and begun to illustrate how it would be useful to apply Moral Foundations Theory to big data. We created a pilot corpus of NYT articles from January 2015, calculated corpus statistics for the data, and analyzed these preliminary findings. We found that the Care/Harm and Authority/Subversion foundations were the most used in both comments and articles. We also found that when we normalized the moral foundation word use, commenters and articles both had very similar mean scores. These findings are just the beginning of what can be learned when we apply Moral Foundation Theory to machine learning and computational linguistics.

Acknowledgements

We would like to acknowledge the support and advice of Professor Muhammad Abdul-Mageed and Professor Patrick Shih. We would also like to thank Dana Skold for his assistance with brainstorming ideas for this project. Finally, we are indebted to all of the students of ILS-Z639 for offering such a great community in which to work.

References

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2(1): 1-135. (Chapters 4 & 5).

Daniel Kahneman. 2011. *Thinking, fast and slow*. Macmillan.

Eyal Sagi and Morteza Dehghani. 2013). "Measuring Moral Rhetoric in Text." *Social Science Computer Review 32*(2): 132-144.

Jeremy A. Frimer, Jeremy C. Biesanz, Lawrence J. Walker, & Callan W. MacKinlay. 2013. "Liberals and Conservatives Rely on Common Moral Foundations When Making Moral

Judgments About Influential People." *Journal of Personality and Social Psychology* **104**(6): 1040-1059.

Jesse Graham, Jonathan Haidt, & Brian A. Nosek. 2009. "Liberals and conservatives rely on different sets of moral foundations." *Journal of Personality and Social Psychology* 96(5): 1029 –1046.

Jesse Graham, Jonathan Haidt, Sena Koleva, Matt Motyl, Ravi Iyer, Sean P. Wojcik & Peter H. Ditto. (2013). *Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism*, Academic Press.

Jonathan Haidt. 2012. *The righteous mind: Why good people are divided by politics and religion*. Vintage.

Jonathan Haidt & Craig Joseph. 2004. Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus, 133*, 55–66.

Jonathan Haidt & Craig Joseph. 2008. The Moral Mind: How Five Sets of Innate Intuitions Guide the Development of Many Culture-Specific Virtues, and Perhaps Even Modules. *The Innate Mind, Volume 3: Foundations and the Future*. P. Carruthers;, S. Laurence; and S. Stich, Oxford University Press.

Jonathan Haidt, & Jesse Graham. (2007). "When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize." *Social Justice Research* 20(1): 98-116.

Morteza Dehghani, Kenji Sagae, Sonya Sachdeva & Jonathan Gratch. 2014. Analyzing Political Rhetoric in Conservative and Liberal Weblogs Related to the Construction of the "Ground Zero Mosque", *Journal of Information Technology & Politics*, *11*(1), 1-14

Paul Ekman. (1992). An argument for basic emotions. *Cognition & emotion*, *6*(3-4): 169-200.

Ravi Iyer, Spassena Koleva, Jesse Graham, Peter Ditto & Jonathan Haidt. 2012. Understanding libertarian morality: The psychological dispositions of self-identified libertarians. *PloS one*, *7*(8)

Saima Aman and Stan Szpakowicz. 2007. Identifying expressions of emotion in text. In Vclav Matouˇsek and Pavel Mautner, editors, *Text, Speech and Dialogue*, volume 4629 of Lecture Notes in Computer Science, pages 196-205. Springer Berlin / Heidelberg.

Stephen Vaisey & Andrew Miles. 2014. "Tools from moral psychology for measuring personal moral culture." *Theory and Society(43)*: 311-–332.

The Authors

Ashley Dainas

Ashley Dainas is a first year PhD student in the department of Information Science at Indiana University, Bloomington. She is interested in Computer-Mediated Communication (CMC) with a particular focus on multimodal CMC and internet memes. Her current work examines the use of reaction GIFs which are image files inserted into social media conversations.

Vipul Munot

Vipul Munot is Strategic, multidisciplinary enthusiast & vivid explorer with an eye for innovation and pixel perfection. My skill set is vast, my greatest expertise revolve in the worlds of interactive design, social media, business analysis and business intelligence. My wish is to combine my knowledge and experience in these areas, to deliver the best creative to my employer's clients and their audiences. I love exploring new things, travelling, trekking, music and coffee.

Satoshi Tsutsui

Satoshi is a first year Ph.D. student in information science at Indiana University, Bloomington. He just got a bachelor's degree in computer science at 2015 from Keio University, Japan. He is interested in almost every domain in artificial intelligence and data mining such as computer vision, natural language processing, semantic web, machine learning including deep learning, graph mining and others.

Appendix I: Charts and Graphs

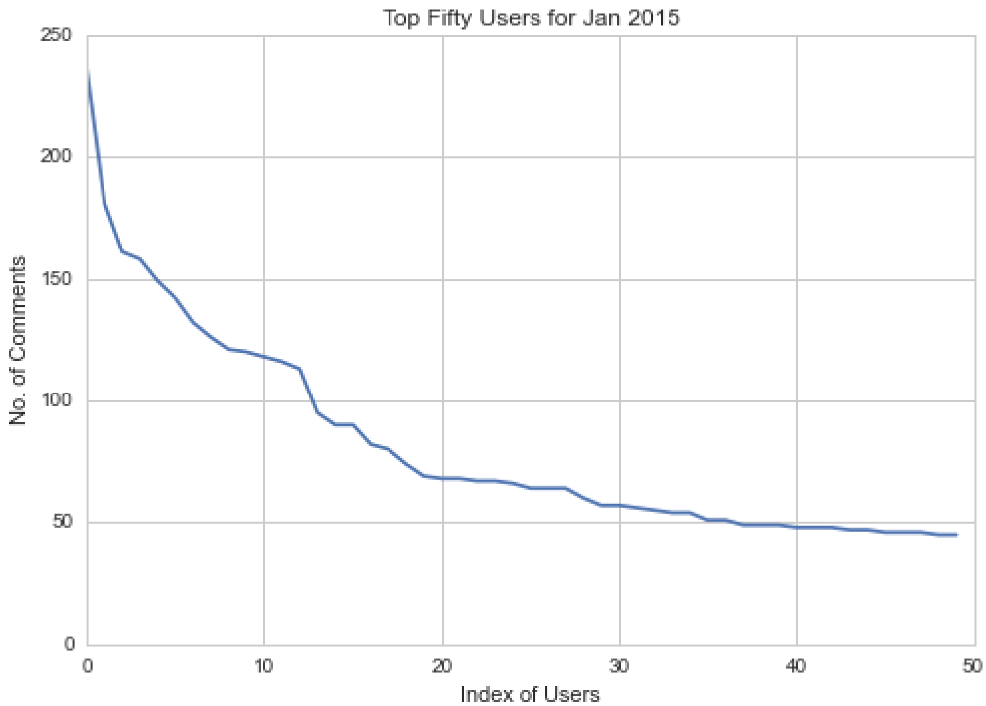

Figure 1: The number of comments by the top fifty users for January 2015.

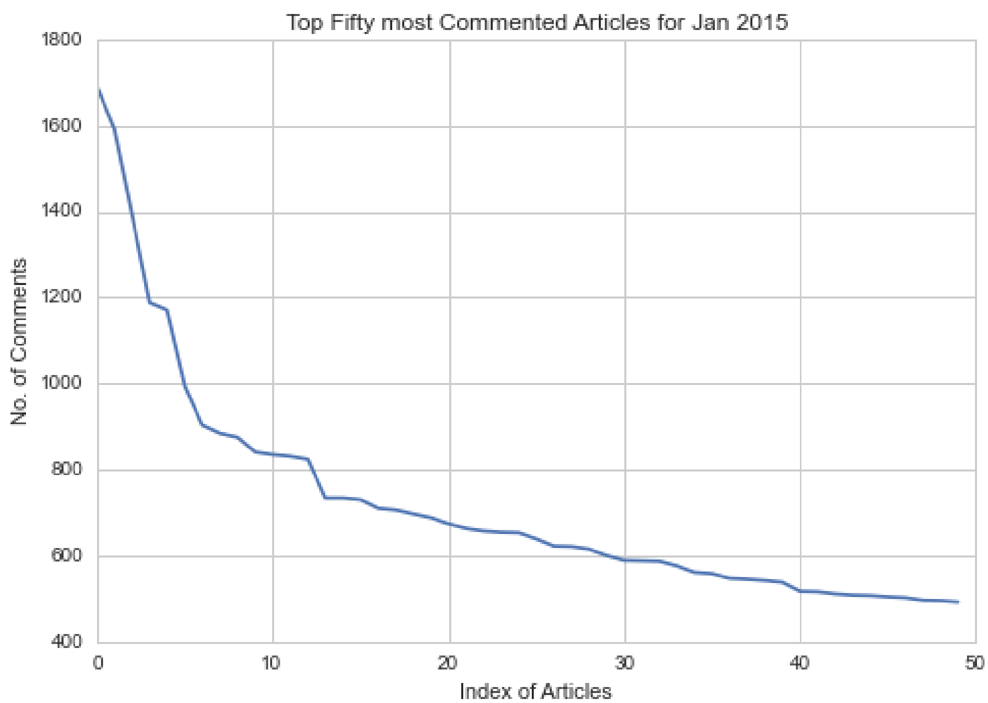

Figure 2: The articles from January 2015 with the highest number of comments.

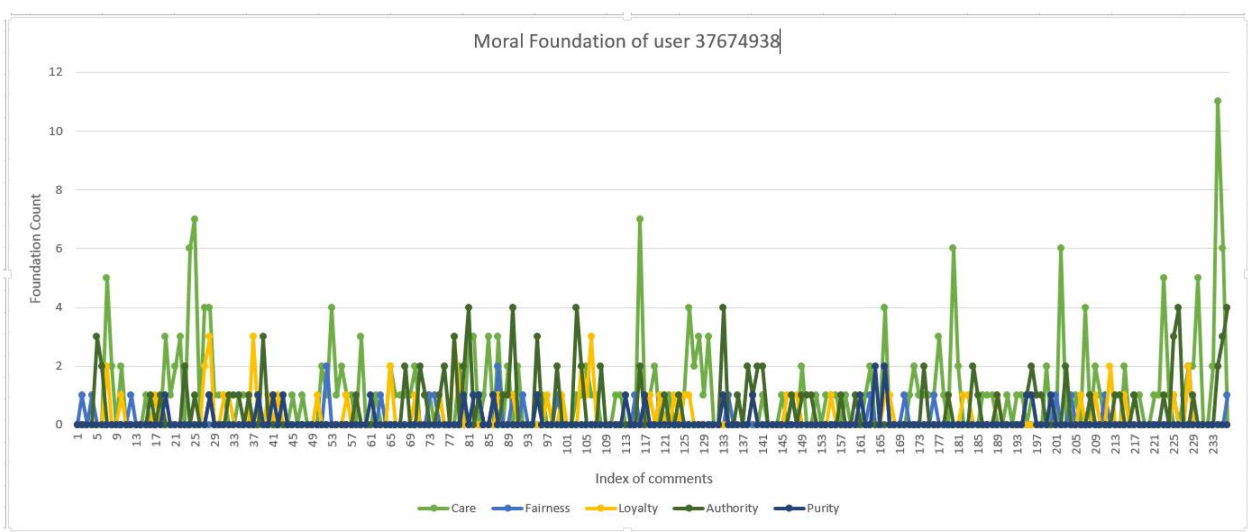

Figure 3: Moral Foundations of user 37674938. This user made 236 comments in January 2015. This was the most comments made by a user that month.

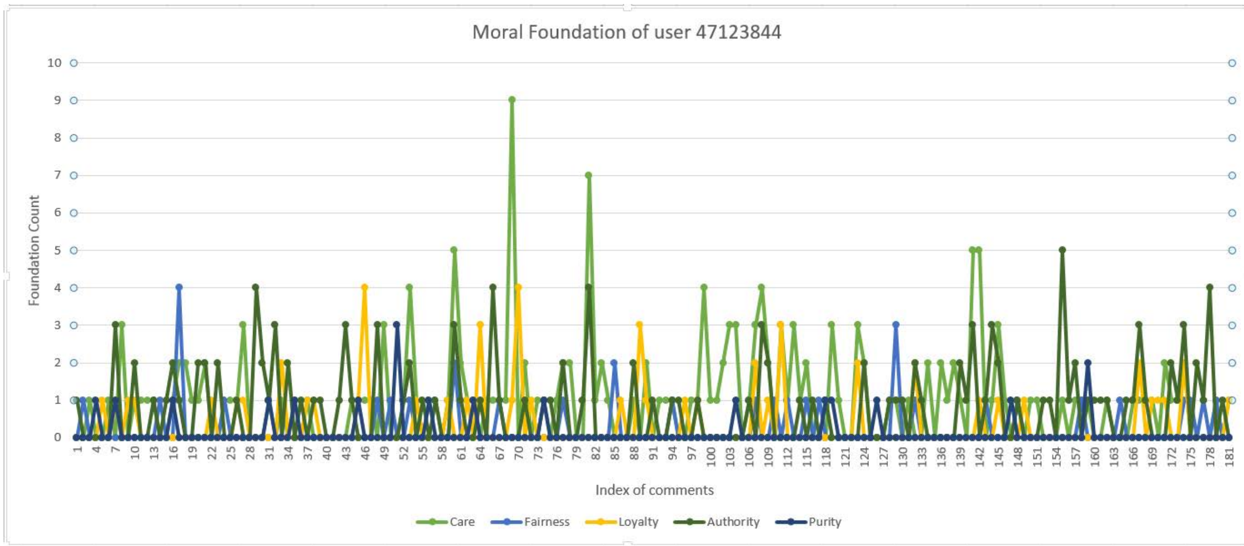

Figure 4: Moral Foundations of user 4123844. This user made 181 comments, which is the second most number of comments in January 2015.

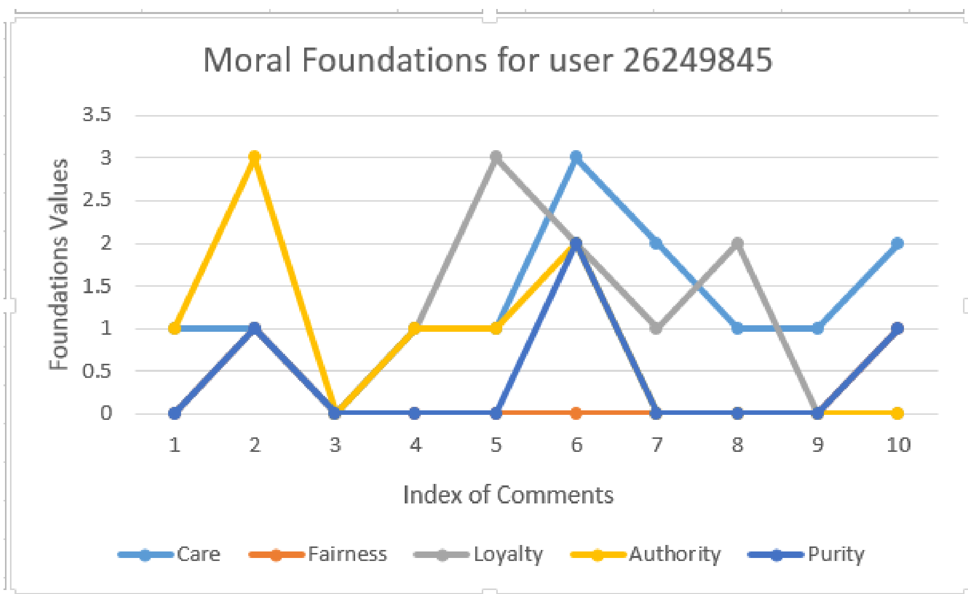

Figure 5: Moral Foundations of user 26249845. This is one of the more casual NYT commenters, who only made 10 comments in January 2015.

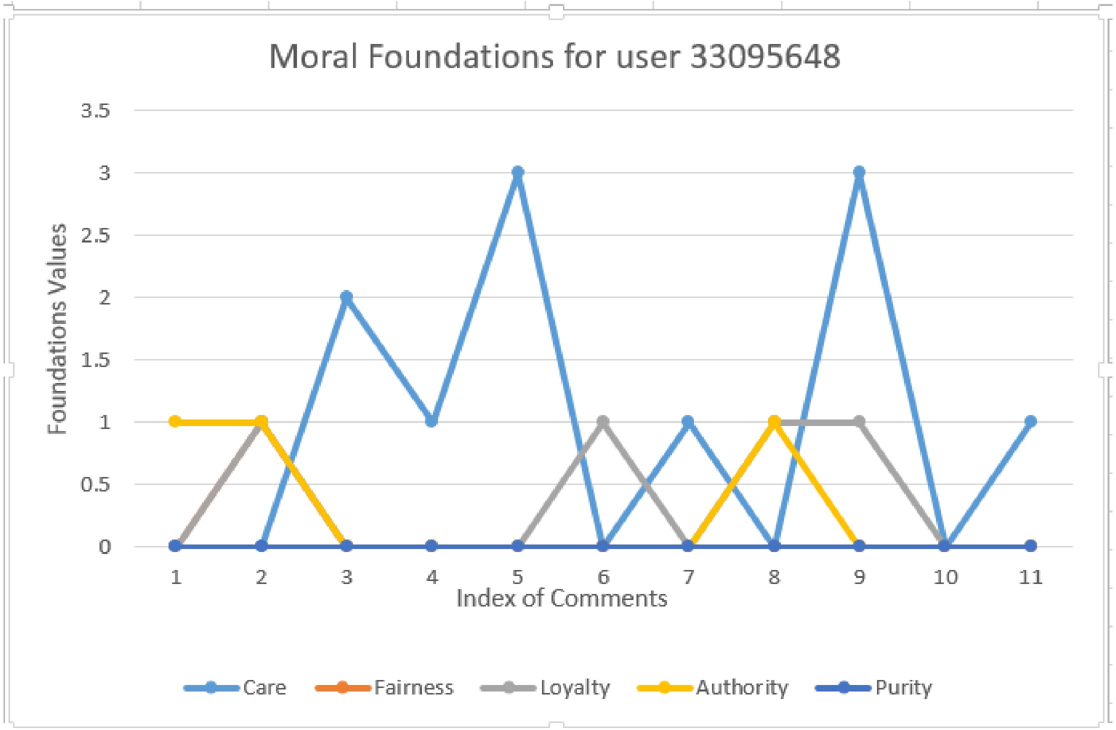

Figure 6: Moral Foundations of user 33095648. This is one of the more casual NYT commenters, who only made 11 comments in January 2015.

Appendix II: Personal Contributions and Absences

Ashley Dainas

1. Came up with the idea
2. Recruited Satoshi
3. Wrote up the proposal
4. Ran the group meetings, at least the parts that did not involve deeply technical discussion
5. Figured out that getting data from social media wouldn't work
6. Chose the New York Times as one of four options to pull data from
7. Analyzed one article and 60 comments by hand so that Satoshi and Vipul has something to test their code against
8. Did the hand annotation counts for each moral foundation, the word count and decided what kind of analysis I thought we should do, though the actual analysis was done by Satoshi and Vipul
9. Did the review of the literature
10. Wrote the entire PowerPoint
11. Figured out what parts of the data were meaningful
12. Figured out which graphs were actually helpful and helped figure out how to convey information meaningfully
13. Realized the need to normalize the data
14. Drew conclusions from the data
15. Came up with future ideas and directions
16. Wrote the abstract, introduction, literature review, data, methods, analysis, discussion and conclusions
17. Helped Vipul write the results section entails
18. Rewrote the results section
19. Formatted the paper, the tables charts and the citations

I missed two days of class, one for Rosh Hashanah on September 14 and got permission to do so. I also missed class on November 30 and I did not get permission for this absence.

Vipul Munot

1. Finding possible practical approach to implement Moral Foundation Theory.
2. Basic research for fetching data from NYT.
3. I used NYT API to fetch URL's and basic data such as article highlights, comments, etc.
4. I then concentrated on analysis of the raw data. I used K-means clustering to get some results. But, it didn't gave any significant results.
5. Then, I cleaned the data removing articles with zero comments.
6. Then I used data to basic corpus statistics to find out mean, max, median, mode for both articles and comments.
7. Then I calculated the maximum comments for the dataset and power users for the month January 2015.
8. I tried implementing graphs for each dataset but it didn't worked as there were too many data points to be plotted.
9. I calculated articles analysis, comments analysis and user analysis sections.

Missed one class.

Satoshi Tsutsui

1. I developed program to retrieve article and its comments from NYT using official API and HTML scraping.
2. I developed a program to count up moral foundation words using stemmers from NLTK and StanfordNLP.

Is not in the class and therefore could not miss any classes.