Understanding Rating Behavior based on Moral Foundations: The case of Yelp Reviews

Pegah Nokhiz
Department of EECS
The University of Kansas
Lawrence, Kansas
Email: pnokhiz@ku.edu

Fengjun Li
Department of EECS
The University of Kansas
Lawrence, Kansas
Email: fli@ku.edu

***Abstract*— Moral inclinations expressed in user-generated content such as online reviews can provide useful insight to understand and predict people's rating behavior. In this work, we extracted a corpus of over 7,000 online reviews on Yelp that express moral concerns, and associated them to five moral factors defined in Moral Foundations Theory using the Doc2Vec natural language processing technique. We compared the rating distributions between the regular reviewers and the moral-concerned reviewers, and found that their rating patterns significantly differ from each other. Our findings also indicate that people with moral concerns tend to rate lower if a moral foundation is violated. Moreover, among the five moral factors, purity is the most distinctive moral foundation.**

***Keywords*- Moral Foundations Dictionary; Social Networks; Paragraph Vector; Natural Language Processing**

I. INTRODUCTION

Nowadays people's interactions in social networks form the backbone of many daily activities. These interactions are considered a new source that reveals human's psychological and behavioral patterns. The influences on users' online activities has been extensively studied in recent years in the communities of computer science, sociology, management and psychology. Besides the textual content of reviews, several variables, such as the counts of upvotes and downvotes, usefulness, coolness, etc., are introduced into the online review platform to provide useful information about the perceived quality or trustworthiness of the reviews and/or reviewers.

Among several factors that may affect one's social behavior, moral inclinations, which were the first insights in intellectual history [9], play an important role in one's attitude and social interactions with others. According to the sociologist Christian Smith "humans are moral, believing, narrating animals" [2, 9]. However, there is little work studying from the moral aspect of the online reviews.

Morality coexists within cultural values and psychological inclinations. With a "binding" approach towards morality, moral systems bind people together by concepts such as family, group, and nation, and *ingroup* in general. For example, some moral systems value groups above individuals and consider suppressing individual desires as virtues such as *purity* and *authority*. Other moral systems disregard groups but emphasize the "individuals" welfare by employing moral foundations such as *harm* and *fairness* [9]. *Purity* is also the main basis for religious laws and the main morality virtue to distinguish moral boundaries [18]. Such moral values represent people's emotions. If a person is inclined to a specific moral virtue, they will feel glad if that moral foundation is practiced or supported [10]. Otherwise, they will feel anger and contempt if a moral virtue is disregarded [19].

Moral inclinations have been studied to distinguish political parties based on the concepts each party tends to endorse. For example, while liberals tend to endorse *harm* and *fairness*, conservatives believe in all moral virtues with less emphasis on *harm* and *fairness* [7, 11]. Similarly, moral values and individuals' moral inclinations influence their expression of opinions, for example in terms of ratings in online reviews.

There have been several studies that model social phenomena based on the insight provided by the moral values extracted from the user-generated content. [1] modelled ideological tendencies by highlighting morally sensitive issues such as same-sex marriage. In [8], moral sentiments were obtained using word embedding methods and the moral rhetoric over time was extracted to examine the evolution of the moral tendencies. [21] studied morally sensitive datasets and the moral loadings of vice keywords in daily tweets. Some studies employed this technique to associate people's social distances with their moral loadings [15]. However, none of these studies quantify moral loadings for online reviews, nor the way they influence people's rating behavior when any immoral practice is involved.

We believe investigating the relationships between a reviewer's moral inclinations and her rating behavior can help understanding the moral, psychological, and cultural intricacies of human nature that potentially affect their online review activities. In this way, we can simplify some of the complications of human interactions by analyzing the moral concerns involved in these interactions.

In this work, we study reviews on Yelp.com, which is a popular online social platform for rating businesses, to investigate people's rating patterns in online reviews as well as how individuals' moral inclinations affect their ratings. In particular, we are interested in three research problems: (1) if the reviewers' ratings change in the face of moral violations and how this change manifests itself for each moral foundation. (2) If morally-inclined reviewers tend to elicit the same tendencies in their general average rating. And (3) if the moral loading of the user is an important factor to study their average rating behavior.

To answer these questions, we apply Doc2Vec, a Natural Language Processing (NLP) technique and *Moral Foundations Theory* (MFT) [11, 12], a leading conceptual framework in moral psychology. As defined by MFT, a given text can be moral if it contains one or more moral values or non-moral. Using Doc2Vec, we analyze the semantic and syntactical meaning of textual content, and identify reviews with morality (or immorality) associated content. For moral-related reviews, we associate it with each of the five moral foundations to calculate its moral loadings. In this way, we can understand the moral concerns expressed in a review and quantify the moral inclinations of the reviewer.

II. Dataset and Data preprocessing

Yelp.com is a popular online social network for rating businesses. In this work, we use an open dataset from the Yelp Dataset Challenge Round 10 [30], which includes 4,153,151 reviews on various kinds of businesses with each review being rated from 1 to 5 stars.

In our study, because we concern about the moral inclinations of the reviewers, we first constructed a morality-relevant dataset of 7,039 English reviews by filtering the reviews with keywords "moral" and "ethic". In our data preprocessing, we removed all extra white spaces and punctuations in the text and converted the capital letters to small letters to avoid extra preprocessing for the uppercase letters.

III. Background

A. Moral Foundations Theory

Individuals hold their own moral values to determine right and wrong, however, the definition of moral or immoral vary widely due to contextual and cultural differences. To understand why morality varies across cultures and extract the similarities, MFT explains morality varies as a function of five moral factors, namely *moral foundations*: (1) *Harm* as disliking others' pain; (2) *Fairness* as doing justice based on common rules; (3) *Ingroup* as being loyal to one's family or nation; (4) *Authority* as respecting and obeying rules and traditions; and (5) *Purity* as feeling aversion towards repulsive things [21].

In MFT, a dictionary consisting of keywords and their stems related to the five moral categories, known as the *Moral Foundations dictionary (MFD)* [29], is proposed by Haidt and Graham to represent each moral foundation with a set of keywords [7, 12]. MFD divides each category into vice and virtue keywords. Virtue keywords support their corresponding moral foundation, e.g. "shelter" or "protect" for the *harm* virtue. Similarly, vice keywords incorporate the words that violate the moral virtue, e.g. "suffer" or "hurt" for the *harm* vice. In this work, we use 149 vice keywords in MFD for our analysis.

B. Doc2Vec

Word2Vec is a word-embedding method of natural language processing recently developed by Google [20]. It is a two-layer neural network to vectorize the words based on the given text context. Word2Vec performs a skip-gram and bag-of-words approach to do the word embeddings. It returns the words and their corresponding vectors in the semantic space, in which similar words are closer to each other [25, 26, 28].

Doc2Vec is an extension to Word2Vec, which improves Word2Vec by enabling representation of paragraphs and longer blocks of text as individual vectors. Besides the word vectors, a new paragraph vector is defined for every paragraph. To quantify the moral loadings, we adopted the Doc2Vec method, and used its Python's genism module [23] to learn paragraph vectors, which is the continuous distributed vector of representations for pieces of texts [20].

Similar to Word2Vec, Doc2Vec has two versions – the distributed bag-of-words paragraph vectors model (i.e., PV-DBOW) model and the distributed memory paragraph vectors model (i.e., PV-DM). PV-DBOW is similar to the Skip-gram model in word vectors, however, it replaces the input by a specific paragraph token that symbolizes the documents. Unlike PV-DBOW that ignores the order of the words, PV-DM take the word order in a small context into account so that important information of a paragraph is preserved. The paragraph token acts as a memory of the context, which is sampled from a sliding window over the paragraph. The paragraph vector can be constructed as either the concatenation or the average of the words in the context, known as the Distributed Memory Paragraph Vector model with concatenated (DMC) or averaged (DMM) paragraph vectors, respectively.

It has been shown that the PV-DM model performs better than the PV-DBOW model because the latter ignores the context words by directly using random initialized words sampled from paragraphs [20]. Therefore, we adopt the PV-DM model as the word-embedding method in this work.

C. Semantic Similarity to Moral Foundations

The Doc2Vec model learns paragraph vectors from unlabeled text data of a variable length, which makes it an attractive method to process the textual content of online reviews in this work. Therefore, we adopted the PV-DM model to convert each online review as a document to a vector in the semantic space. In particular, we implemented the DMM approach and utilized the average word vectors of the key words of each moral foundation. Meanwhile, words in the reviews are represented as vectors in a vector space where semantically similar words have similar vector representations. In this way, we can calculate the text similarity of the review to a moral foundation as the cosine similarity of the document vectors and MF words by averaging MF words' respective vectors. As a result, each reviewer in the system will have five *moral loadings*, which are the average of the reviewers' cosine similarities in each MF, respectively.

IV. Our method

In Word2Vec, the aim is to predict a word given its surrounding words. Given a neural network of only one hidden layer, the input IDs are the context words which are the words surrounding the output word. The output layer is the word of interest for prediction. The neural network tries to learn and adjust the corresponding weights by performing the training process to maximize the probability of the output word. These weights will be the vectorized representation of the words after several rounds of training. Doc2Vec follows the same pattern; however, it has additional nodes as special tokens to symbolize each document. If we represent the feature that symbolizes the document contexts as $\boldsymbol{D}$ and the context words as $\boldsymbol{W}$ which are the words in a window surrounding the output word, Doc2Vec's goal is to maximize the following log probability:

$$\max \sum\nolimits_{\forall(output\ word,\ W,\ D)} \log P(output\ word|W, D) \qquad (1)$$

This stage provides us with the document embeddings and the word embeddings of the training corpus. The second stage is "the inference stage" for the documents that we have not seen yet. This process is similar to the previous maximization step. However, in this stage we can keep the weights as constants and then learn $\boldsymbol{D}$ for the testing corpus [20].

We considered each review as a document. We tokenized the review corpus based on the whitespace and removed non-alphanumeric characters and the less frequently occurred tokens. Then, we treated each review as a separate paragraph to train the PV-DM model. We embedded the documents into vectors of size 100 in the semantic space. In addition, we used a window size of 10 and negative sampling of size 5 which indicates the count of the noise words drawn by negative sampling.

After training, we obtained the vectors representation for each token in the corpus. For each sentence, we formed a vector, $\mathbf{r} = (\mathrm{r}_1, \dots, \mathrm{r}_v)^{\mathrm{T}}$, corresponding to the summation of the vectors of all the tokens in the sentence. Similarly, for each of the five moral foundations, we formed a vector, $\mathbf{f} = (\mathrm{f}_1, \dots, \mathrm{f}_v)^{\mathrm{T}}$, corresponding to the summation of the vectors of all the vice keywords in that moral foundation.

We used the cosine similarity between a review vector $\mathbf{r}$ and a moral foundation vector $\mathbf{f}$ as a measure for the similarity of a review to a moral foundation, which is calculated as the dot product between the two vectors normalized by their norms. In the moral context, a cosine similarity close to 1 indicates that the review is semantically similar to that moral foundation, while a cosine similarity close to 0 indicates that the document and the molar foundation are not semantically related. Therefore, we define the *moral loading* m_{ij} for a review as cosine similarity between the moral foundation f_i, where $i \in \{1, 2, 3, 4, 5\}$ is an index representing each moral foundation, and the review r_j. Similarly, we define the moral loading M_{ij} for a reviewer u_j as the average of the moral loadings of all his reviews.

V. Study I

In this work, we conducted two experimental studies to explore the relationship between people's moral concerns and their rating behavior. In particular, we first identified the frequency of the ratings in each moral category and calculated the conditional relative frequency considering the unbalanced datasets. Secondly, we tracked the regular users who have rated the same businesses as the moral-concerned users identified in our moral corpora, and studied the differences in their rating behaviors.

A. Identify Relevant Users based on the Moral Loadings

In the first study, we aim to investigate if people who care more about morality will rate differently from the regular users who do not show a clear moral inclination in the face of moral violations. We are also interested in exploring the different ways that the reviewers rate under different morality contexts.

To tackle this problem, we first need to identify individuals associated with each of the five moral foundations, i.e., *harm*, *fairness*, *ingroup*, *authority*, and *purity*. As described in Section IV, we calculated the moral loadings in each MF category as the cosine similarity for a review and the keywords in that MF category. We then ranked the reviews based on their moral loadings.

To locate the most similar document to each moral foundation, we defined a cosine similarity threshold. It is pointed out in [17] that the threshold for the cosine similarity measures in document comparison should be dynamically adjusted, since low cosine thresholds can produce good results in terms of precision and recall. In fact, setting a too high threshold without considering the specific context's experimental results will result in excluding documents that are similar. As recommended in [16], "researchers taking the factor analysis approach to LSA should not apply 0.40 or some similarly preset loading threshold, but instead apply an empirically derived threshold, validated by a domain expert because thresholds as low as 0.18 were found acceptable." Following this idea, we experimentally set the threshold for cosine similarity in our moral corpora to 0.2. Our empirical analysis showed this threshold as a good boundary to distinguish morally similar documents. In the morally filtered dataset of 7,039 reviews, there are 5,782 reviews with the cosine similarity larger than 0.2.

In particular, we had 1,002 reviews for the *Ingroup* MF category, 1,115 reviews for *authority*, 1,118 reviews for *harm*, 1,188 for reviews *fairness*, and 1,359 reviews for *purity*. In each case, the reviews with a loading above 0.2 maintained a reasonably strong relevance to the respective moral foundation category. In the meantime, the result provides a reasonably large size of morally relevant reviews to be used in further analysis.

B. Relationship between Users' Moral Concerns and Ratings

Next, we studied the direct relationship between a user's own moral inclination and her rating behavior. In the Yelp dataset, for each review, a user explicitly gives a star rating, ranging from 1 star to 5 stars. The result is shown in Figure 1. For each moral foundation category, we show the distribution of reviews in different star ratings. In all moral foundation categories, it is obvious that the reviews with 1-star rating outnumber the reviews of any other rating. This indicates users who care more about the moral concerns tend to give low (i.e., 1-star) ratings.

However, when we studied the original dataset of 4,153,151 reviews, we found that the dataset has an unbalanced number ratings. The number of reviews rated in 5 stars is much larger than the reviews rated in 1 star. In particular, there are 1,704,200 reviews (41%) of 5-star rating and 540,377 reviews (13%) of 1-star rating.

Due to this imbalance, we believe the conditional distribution should be a more informative and reasonable than the direct distribution. Therefore, we defined the *conditional relative frequency* of each rating as:

$$\textit{Conditional relative frequency} = \frac{\#\textit{rating in each MF}}{\#\textit{rating in the dataset of interest}} \qquad (2)$$

where the dataset of interest can be the entire Yelp dataset or the morally filtered dataset of 7,039 reviews.

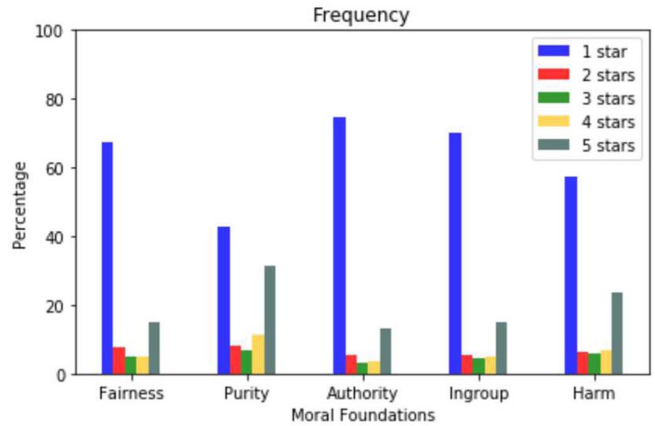

Figure 1. Frequency of each rating in five moral corpora

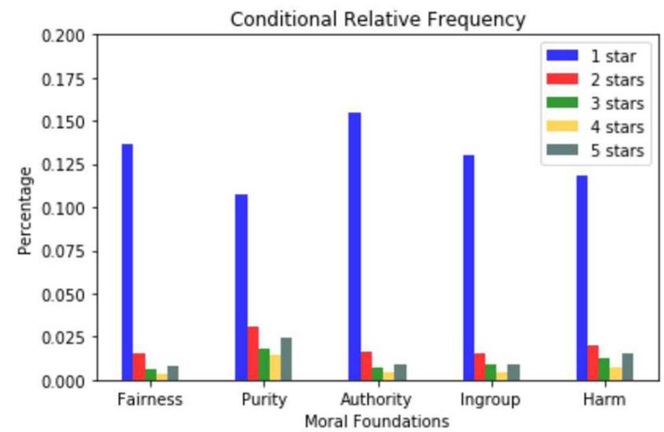

Figure 2. Conditional relative frequency of each rating relative to the dataset of 4,153,151 reviews

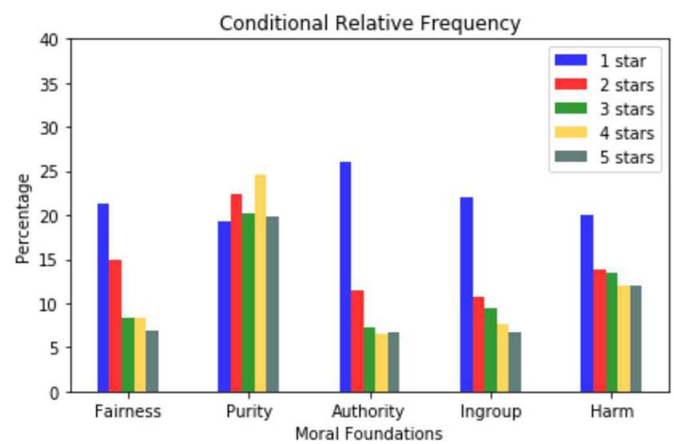

Figure 3. Conditional relative frequency of each rating relative to the dataset of 7,039 reviews

We first analyzed the conditional relative frequency in relevance to the entire dataset with 4,153,151 reviews. As shown in Figure 2, in all five moral foundation categories, the frequency of reviews in 1-star rating is significantly larger than the number of reviews in other star ratings.

Next, we calculated the conditional relative frequency in relevance to our moral dataset (i.e., the dataset with 7,039 moral-relevant reviews). The result is shown in Figure 3. For instance, in the moral dataset of 7,039 reviews, there are more than 20% reviews in 1-star rating considered related to *fairness*, while only 6% reviews in 5-star rating are considered related to *fairness*.

Moreover, in all moral foundation categories except the *purity* category, a consistent stepwise decreasing pattern was observed in the conditional relative distribution of reviews with different star ratings. This indicates that users giving lower ratings tend to consider more about the *fairness*, *authority*, *ingroup*, and *harm* aspects in their reviews, while users giving higher ratings have less considerations in mind.

The only exception is the *purity* category, in which no matter which star rating is given, an approximately same relative percentage of users care about *purity* (e.g., “disgust”, “gross”, “indecent”, “trashy”, etc.) in the reviews. This finding is in line with some previous work on moral foundations. For example, Dehghani et al. [15] investigated the influence of *purity* homophily as a predictor of social distances. Their results indicated that comparing with other moral foundations, *purity* is the main predictor of the social distances.

C. Users’ Rating Behavior

In this task, we studied the rating behavior of regular users and users with moral inclinations.

In *task I-A*, we identified a set of users whose reviews are related to five moral foundations. We also searched the entire Yelp dataset to locate another group of 370,221 users, who had reviewed the same set of businesses that the moral-concerned users reviewed. Therefore, we constructed two user sets, i.e., regular users and moral-concerned users.

We first show regular users’ rating distribution in Figure 4 (left). For the target set of businesses, this figure shows percentages of reviews with different ratings. Overall, there are more reviews with 4 and 5 star ratings than the ones with 1-3 star ratings. We also plot the conditional relative frequency of review ratings in relevance to the count of each rating in the entire dataset, as shown in Figure 4 (right). This indicates the rating behavior of the selected regular user set is compatible to the general users and the selected regular users are not biased.

Next, we studied the rating distribution of the users in the moral set. The frequency and conditional relative frequency of ratings of moral-concerned users are shown in Figure 5. In both plots, there are significantly more reviews with 1-star rating than reviews with higher ratings. This is in line with our findings in the previous task.

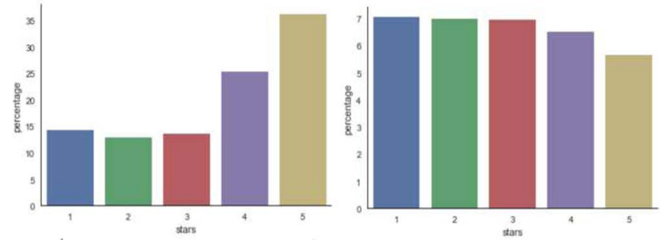

Figure 4. Frequency (left) and conditional relative frequency (right) of the ratings of regular users

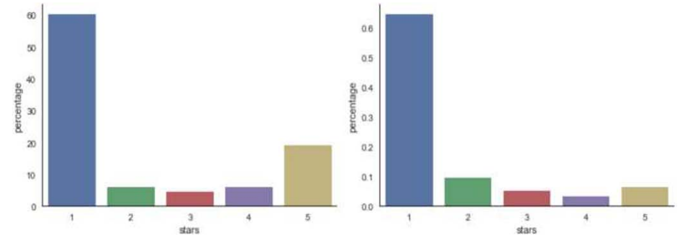

Figure 5. Frequency (left) and conditional relative frequency (right) of the ratings of moral-concerned users

TABLE I. CORRELATIONS BETWEEN STARS’COUNT AND COSINE SIMILARITIES

Moral Foundations	Correlation
Fairness	-0.302
Harm	-0.145
Autority	**-0.350**
Ingroup	-0.284
Purity	0.065

Comparing Figure 4 and Figure 5, we clearly see that regarding the same set of businesses, regular users who do not care about the moral foundations or face any moral violations, rate differently from users who do have moral concerns. In particular, people who care about moral violations tend to rate lower than regular users. Moreover, the moral-concerned users tend to give the lowest rating compared to the regular users.

We further studied the correlation between the moral loadings and the count of star ratings for our five moral corpora. As shown in Table I, the negative correlations for *ingroup*, *fairness*, *harm*, and *authority* are compatible with the previous bar plots, since the higher ratings the reviews have the smaller their moral loadings will be.

It is worth to point out that the correlation result of the *purity* category also went along with the unbalanced rating pattern of the reviewers in the *purity* moral corpus.

VI. STUDY II

A. User's Average Rating Behavior

In the previous study, we showed that users with moral considerations rate differently from the regular users. In this study, we aim to examine the rating behavior of the moral-concerned users by comparing the average ratings of their moral-related reviews and the other reviews that do not show clear moral relevance. In other words, if the users with high moral loadings show the same moral inclinations in their general average ratings. For each user in the moral set identified in *Task I-B*, we calculated the rating difference of the user as:

$$d_{ij} = |r_{ij} - g_j|, \qquad 0 \leq d_{ij} < 4 \qquad (3)$$

where g_i denotes the average rating of all her reviews, and r_{ij} denotes the average rating of her reviews regarding the moral foundation i and j is an index for a reviewer u_j .

For each of the five moral foundations, we calculated the *rating difference* for all the users with reviews relevant to this moral foundation. The corresponding cumulative density function is shown in Figure 6 to provide direct statistical insights about the moral-relevant users' rating behavior. Generally speaking, these reviewers tend to show the same rating behavior in their overall reviews as compared to their moral-concerned reviews. As shown in Figure 6, the maximal of the average rating difference is 3.8, and more than 50% of users have a rating difference smaller than 1.5 stars, which is close to the theoretic average rating difference of 2 stars.

Next, we define the weighted average rating difference by incorporating the moral loading of each user as the weight. This is because users related to one moral foundation have different moral loadings, which indicates the degree of inclination to the moral foundations. Consequently, we calculate the weighted average rating difference as $M_{ij} \times |r_{ij} - g_i|$, and show the corresponding cumulative density function results for five moral foundations in Figure 7.

As shown in Figure 7, over 90% of users have a rating difference smaller than 2.5, and over 20% of users have a rating difference smaller than 1, in all five moral foundations. This indicates that moral-concerned user rates consistently in their moral-concerned reviews and the general reviews.

B. Correlation between Moral Foundations

We also studied the correlations between moral foundations. As shown in Table II, *authority* and *fairness* have the highest correlation, and *purity* and *ingroup* have the lowest value. In fact, *purity* is not highly correlated with any other moral foundation. This is in line with previous studies [15, 18, 21], our findings in Table I, and the unbalanced ratings in study I, which indicates that *purity* is the most peculiar moral foundation.

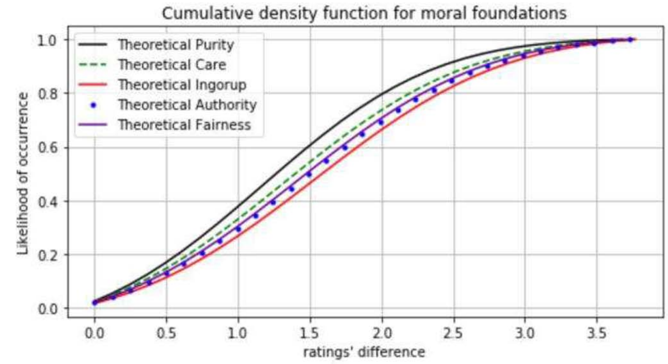

Figure 6. CDF of the absolute difference of average moral ratings and general average moral ratings of the reviwers for each moral foundation

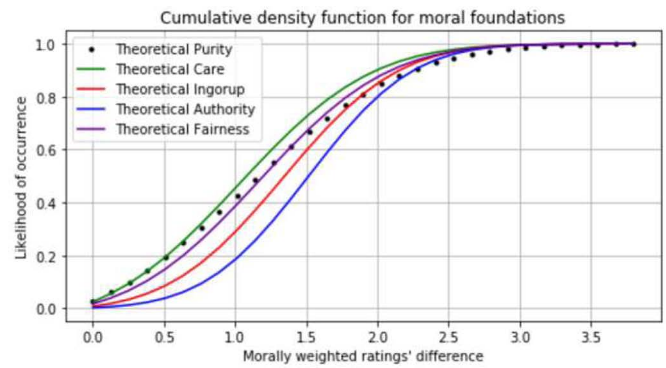

Figure 7. CDF of the morally weighted absolute difference of average moral ratings and general average moral ratings of the reviwers for each moral foundation

TABLE II. COSINE SIMILARITIES' CORRELATIONS

Moral Foundations	Fairness	Harm	Authority	Ingroup	Purity
Fairness	-	0.355	**0.681**	0.384	0.286
Harm	-	-	0.368	0.385	0.332
Autority	-	-	-	0.527	0.145
Ingroup	-	-	-	-	**0.132**
Purity	-	-	-	-	-

We also observed that *ingroup* and *authority* are highly correlated. This may be because they are from the binding foundations [9]. We expected a higher correlation between *harm* and *fairness* since they are both individualizing foundations, however, this was not observed in our results.

Finally, we compute a word cloud with all vice keywords identified in our moral corpora. As shown in Figure 8, keywords refuse and favoritism are highlighted as the most frequent word in our moral corpora.

Figure 8. Word cloud of MFD's vice keywords in our moral corpora

VII. Related Work

This work is related to two research areas. One is the Moral Foundations Theory and its applications to the study of social activities, and the other is the word embedding models and its applications to discover the relationship between a textual input and a given topic of interest.

Moral foundations' effect on social networks has been studied extensively in the previous studies. Dehghani et al. [15] investigated the influence of *purity* homophily as a main predictor of social distances. In addition, Kaur and Sasahara employed big data analysis and word-embeddings to investigate four morally sensitive issues in twitter [21]. In another study, Zhang and Counts used MFT to extract the ideological patterns and predict potential changes with moral foundations as a predicting factor [1]. Landmann and Hess, investigated if specific emotions are elicited by specific moral foundations [6]. Additionally, Garten et al. employed sentiment analysis on a set of tweets and examined the evolution of the moral rhetoric over time [8]. The political positions of people in twitter and their retweeting behavior in inter-community and intra-community retweets was studied based on MFT variables in [5] by Sagi and Dehghani.

In the series of studies of Sagi and Dehghani, the semantic similarity between keywords of a specific corpus and the MFD keywords are defined as the moral loading for the topic of interest. The moral loadings can then be used as a factor to test various morally-relevant hypotheses [4, 13]. Inspired by MFT, researchers in other communities have used machine learning techniques and NLP to create a user-defined dictionary of words [7, 13]. Furthermore, big data techniques have been used by Boyd et al. to investigate the relationship between moral values and the behavioral patterns in texts [13, 22].

The natural language processing embedding tools also have a rich history of studies. Doc2Vec by Le and Mikolov [20] and Word2Vec by Mikolov et al. [26], discussed in the earlier sections, are two of many embedding techniques. Latent Dirichlet Allocation proposed by Blei, Ng, and Jordan, which is a three-level hierarchical Bayesian model for modelling items on top of a set of topics, is another study in document modeling [3]. Deerwester et al. studied latent semantic analysis which uses a bag-of-words model to represent a corpus by a dimensionality-reduced context matrix [24, 27]. Another study of Pennington, Socher, and Manning proposed "GloVe: Global Vectors for Word Representation", which is another word-embedding method that utilizes dimensionality reduction on the co-occurrence word-context matrix [14].

VIII. Conclusion

In this paper, we argue that our daily activities including our actions in social media are bounded with morality and culture to certain extent. To analyze this relationship, we studied the influence of moral foundations on the reviewers' rating behavior in the immorality context.

In particular, we used the moral foundations proposed in the Moral Foundations Theory (MFT) [11, 12] and the vice keywords defined in the Moral Foundations Dictionary (MFD) to support a natural language processing analysis on a dataset of online reviews from the Yelp Challenge. We adopted the word embedding method, *Doc2Vec*, to convert the reviews to vector representations in the sematic space defined by the reviews. By converting the review and the MFD keywords for each moral foundation into vectors, we can calculate the moral loading of the review in the form of cosine similarity between the vector of the document and the vector of the average vice words. Using an experimentally defined threshold for cosine similarities, we identified a moral corpus and a corresponding moral-concerned user set for each moral foundation.

We investigated the frequency and conditional relative frequency of review ratings for the overall moral corpus and the moral corpora associated with each of the five moral foundations, as well as the rating distributions of the regular users who rated the same set of businesses. The comparison shows that the rating pattern of regular users differs significantly from the one of the moral-concerned users. Moreover, our findings indicate that people with moral concerns tend to rate lower if a moral foundation is violated.

For the moral-concerned users, we also studied differences, in terms of the average rating and the weighted average ratings, between their moral-related reviews and their all reviews with no moral consideration. Results in the cumulative density functions reveal that there is a higher likelihood of a smaller difference if we consider each reviewer's moral loading.

Finally, we examined the correlation between the moral foundations themselves. *Purity* was shown to be the most distinctive moral foundation.

We believe this work represents a new avenue of analysis on moral psychology and online social networks. For future work, we will test our findings with larger corpora and more word embedding methods. We will also consider making a comprehensive dictionary to filter the moral reviews. In addition, we anticipate developing an extended version of

MFD, which incorporates more words in the same medium. We also hope to incorporate emotion detection analysis and sentiment analysis in this study to improve the results.

REFERENCES

[1] A.X. Zhang, and S. Counts, "Modeling ideology and predicting policy change with social media: Case of same-sex marriage.," In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2603-2612). ACM, 2015.

[2] C. Smith, Moral, "believing animals: Human personhood and culture.," Oxford University Press, 2003.

[3] D.M. Blei, A.Y. Ng, and M.I. Jordan, M.I.,"Latent dirichlet allocation," *Journal of machine Learning research*, *3*(Jan), pp.993-1022, 2003.

[4] E. Sagi, and M. Dehghani, "Measuring moral rhetoric in text. Social Science Computer Review," Vol. 32 (2), pp. 132-144, (2014a).

[5] E. Sagi, and M. and Dehghani, "Moral rhetoric in Twitter: A case study of the US Federal Shutdown of 2013," In *Proceedings of the Cognitive Science Society* (Vol. 36, No. 36), 2014.

[6] H. Landmann, and U. Hess, "Testing moral foundation theory: Are specific moral emotions elicited by specific moral transgressions?," *Journal of Moral Education*, pp.1-14, 2017.

[7] J. Graham, J. Haidt, and B.A. Nosek, "Liberals and conservatives rely on different sets of moral foundations," *Journal of personality and social psychology*, *96*(5), p.1029, 2009.

[8] J. Garten, R. Boghrati, J. Hoover, K.M. Johnson, and M. Dehghani, "Morality Between the Lines: Detecting Moral Sentiment In Text," *Proceedings of IJCAI 2016 workshop on Computational Modeling of Attitudes*, New York, NY, 2016.

[9] J. Haidt, "Morality," *Perspectives on psychological science*, *3*(1), pp.65-72, 2008.

[10] J. Haidt, The Positive emotion of elevation, 2000.

[11] J. Haidt, *The righteous mind: Why good people are divided by politics and religion.* Vintage, 2012.

[12] J. Haidt, and C. Joseph, "Intuitive ethics: How innately prepared intuitions generate culturally variable virtues," *Daedalus*, *133*(4), pp.55-66, 2004.

[13] J. Hoover, "Into The Wild: Big Data Analytics in Moral Psychology," *structure*, *7*(3), pp.269-279.

[14] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543), 2014.

[15] M. Dehghani, K. Johnson, J. Hoover, E. Sagi, J. Garten, N.J. Parmar, S. Vaisey, R. Iliev, and J. Graham, "Purity homophily in social networks," *Journal of Experimental Psychology: General*, *145*(3), p.366, 2016.

[16] N. Evangelopoulos, X. Zhang, and V.R. Prybutok, "Latent semantic analysis: five methodological recommendations," *European Journal of Information Systems*, *21*(1), pp.70-86, 2012.

[17] P. Penumatsa, M. Ventura, A.C. Graesser, M. Louwerse, X. Hu, Z. Cai, and D.R. Franceschetti, "The right threshold value: What is the right threshold of cosine measure when using latent semantic analysis for evaluating student answers?," *International Journal on Artificial Intelligence Tools*, *15*(05), pp.767-777, 2006.

[18] P. Rozin, J. Haidt, and C.R. McCauley, "Disgust," *Handbook of emotions,* 1993.

[19] P. Rozin, L. Lowery, S. Imada, and J. Haidt, "The CAD triad hypothesis: a mapping between three moral emotions (contempt, anger, disgust) and three moral codes (community, autonomy, divinity)," *Journal of personality and social psychology*, *76*(4), p.574, 1999.

[20] Q. Le, and T. Mikolov, "Distributed representations of sentences and documents," *Proceedings of the 31st International Conference on Machine Learning (ICML-14)* (pp. 1188-1196), 2014.

[21] R. Kaur, K. and Sasahara, "Quantifying moral foundations from various topics on Twitter conversations," *Big Data (Big Data), 2016 IEEE International Conference on* (pp. 2505-2512), IEEE, 2016.

[22] R.L. Boyd, S.R. Wilson, J.W. Pennebaker, M. Kosinski, D.J. Stillwell, R. Mihalcea, "Values in Words: Using Language to Evaluate and Understand Personal Values," International AAAI Conference on Weblogs and Social Media, 2015. Retrieved from https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10482

[23] R. Rehurek, P. and Sojka, "Software framework for topic modelling with large corpor," *In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 2010.

[24] S. Deerwester, S.T. Dumais, G.W. Furnas, T.K. Landauer, and R. Harshman," Indexing by latent semantic analysis," *Journal of the American society for information science*, *41*(6), p.391, 1990.

[25] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[26] T. Mikolov, I. Sutskever, K. Chen, G.S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *Advances in neural information processing systems* (pp. 3111-3119), 2013.

[27] T.K. Landauer, and S.T. Dumais, S.T., "A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge," *Psychological review*, *104*(2), p.211, 1997.

[28] Y. Goldberg, and O. Levy, "word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method," *arXiv preprint arXiv:1402.3722*, 2014.

[29] http://moralfoundations.org/

[30] https://www.yelp.com/dataset/challenge