An Exploration of Semantic Tendencies in Word of Mouth Business Reviews

David W. Vinson and Rick Dale
Cognitive and Information Sciences
University of California, Merced; Merced, USA

***Abstract*—Explicit customer review ratings mark future business success. One important and well-studied aspect of customer satisfaction is a review's affective—positive or negative—valence. More recently, tools from natural language processing (NLP) applied to reviews show less obvious linguistic differences in review texts dependent on reviewer rating. Consistent with this is previous work using *Linguistic Inquiry and Word Count* (LIWC), showing that language use changes depending on one's current psychological state. Finer-grained analyses of review text focusing on less obvious linguistic categories may provide insight into customer values. In an attempt to explore how the content of a review is related to a review's explicit rating, we analyzed review texts using LIWC. LIWC determines the percentage of review text associated with a variety of different psychologically relevant categories such as social or cognitive words. We explore how certain categories of words relate to review ratings and use a support vector machine to determine how well each category predicts reviewer's review rating. We relate our findings to previous work and speculate that businesses would benefit from the application of various Natural Language Processing tools in attempting to obtain comprehensive insight into customer satisfaction. We end with the connection between this work and theories of language use, for which data sets of customer reviews may be useful for exploring the role of psychological state in determining word choice.**

Keywords—language; natural language processing; corpus analysis; support vector machines; word of mouth

I. Introduction

Research amassed over the past two decades suggests future business demand is largely influenced by open-ended peer-authored written reviews [1, 2]. In fact, positive word of mouth (WOM) reviews are touted as the "missing link" in understanding how customer satisfaction impacts future business [3, 4]. Indeed WOM reviews provide a substantial impact on new customers looking to book hotels [5], flights [6] and other travel plans [1].

WOM reviews can be found on websites like Amazon.com, an e-commerce platform featuring products from companies all over the world, or third party sites that focus primarily on business reviews such as Yelp, Inc. Customers can report their experience through explicit ratings usually from 1 (negative) to 5 (positive) stars and open-ended and informal comments [7, 8]. On the surface customer reviews stand to encourage better customer-business relations; providing insight into customer satisfaction beyond that of repeated business. However, satisfied customers are not always repeat customers, even though explicit review ratings can affect the average revenue of a business by approximately 5-9 percent [9]. WOM reviews may be the key to better customer-business relations, but their exact impact remains elusive [4]. The present work explores the relationship between word usage in WOM reviews pertaining to specific psychological categories and explicit reviewer ratings. Businesses may stand to benefit from understanding reviewer language use as it relates to customer satisfaction. In addition, by studying these linguistic aspects of reviews, we may provide an important interface between theories of language use, and commercial contexts in which language is used. We consider both in this paper, beginning with an analysis of word choice in terms of semantic dimensions, and how these correspond to customer experience.

Previous research shows customer satisfaction impacts WOM reviews that in turn impact future business success [1, 3, 9]. Indeed, when customers are highly satisfied their reviews act as a promotional material for businesses reviewed. Potential customers read reviews and determine if the service or product is worth their business. In addition, those interested in goods or services previously reviewed, often have the opportunity to rate how useful or helpful a review was in providing valuable information to the readers. Interestingly, helpful reviews are often those that are given ratings that fall somewhere between highly positive or negative reviews [2]. This is in contrast to the finding that a greater number of reviews occur at the extremes (highly positive or highly negative), with significantly fewer reviews falling somewhere in the middle [8]. Interestingly, what seems to be considered worth reporting and a helpful report are not always the same thing. This highlights the elusive nature of WOM reviews in providing a rich understanding of what customers find both worth reviewing and helpful. Understanding the subtleties of a WOM review can provide insight into what reviewers and review readers consider important.

Recent research suggests the linguistic structure that comprises the combinations of words within a review is influenced by reviewer satisfaction. Specifically, the information density or complexity of a review text is significantly different depending on its explicit positive or negative review rating [10]. Interestingly certain measures of information show a quadratic relationship such that both highly positive and negative reviews contain information-dense language while less extreme reviews tend toward lower information density. Interesting correlations exist between information density and other studies of WOM reviews. Specifically, previous research shows a similar quadratic relationship for the frequency of reviews [8]. Considering previous research showing review readers find less extreme reviews more helpful [2], readers maybe sensitive the structural components of a reviewer's language use. This possibility stands primarily as an example of how a finer-grained analysis of WOM reviews may provide insight into reviewer ratings and review helpfulness not previously uncovered. Through large-scale NLP analysis, it may also shed light on how language users load evaluative communication with certain information content or semantic dimensions. The result would contribute both to understanding WOM, *and* the nature of language use in a realistic context [10].

Understanding the subtle aspects of WOM reviews and how they relate to explicit ratings can provide businesses with a better understanding of customer satisfaction and predictions of future business success. Recent findings show reviews considered helpful by readers are also more readable [11], while social and personality measures determined through a reviewer's use of language within their reviews can be modeled into a personalized recommender system catered to the interests of the individual [12]. Furthermore, analyses of business reviews and reviewer WOM content have provided predictions of future business demand [9, 13], review helpfulness [11] and reviewer deception [9]. However, no single aspect of WOM reviews has been shown to elucidate the exact impact of customer reports on future business success. More likely, a combination of different linguistic components will collectively bring about a better understanding of how exactly WOM reviews impact business success. Certain tools from an NLP domain that focus on analyzing the content of a text can allow for the exploration of many different psychological and social categories present within review texts simultaneously. One such tool, *Linguistic Inquiry and Word Count* (LIWC) has been fruitfully applied in a wide variety of fields [14]. LIWC analyzes texts at the word level and provides a percentage of the number of words within a text that fit into specific categories such as *cognitive*, *social*, and *affective* (positive or negative) words. Understanding how certain words within a review are associated with review ratings may provide insight into the focus of customer attention.

Stylistic differences at the word level in written and spoken messages reveal important aspects of a speaker's implicit feelings and behavior [12]. For example, patients suffering from mental illness change the style of their language as they become healthier [14, 15]; patients suffering from depression show a greater use of first person singular pronouns than individuals who have never suffered from depression [16]; word usage is dependent on social structure [17, 18]; increased usage of cognitive words lead to benefits of patients after serious trauma [19]; there exists a strong relationship between life expectancy and positive word usage [20]. Analysis of texts in the field of clinical psychology is not new, yet it continues to elucidate a variety of behavioral and clinical behaviors. More recently, there exists a growing interest in the application of linguistic analyses on WOM business reviews [21].

Importantly, some researchers have focused on how specific words within a review, such as ones conveying positive or negative affect can help to predict how many reviews a business will receive in the near future [13][1][2]. These studies show that understanding the general features of a review text are related to explicit behaviors of the reviewer and reader. Yet, focusing on general aspects like that of positive or negative words can leave out rich linguistic content nested within the text. Such minutiae may elucidate what psychological processes underlie a reviewer's explicit ratings. Specifically, words that are not outright positive or negative, but belong to different psychological categories, such as social or cognitive words, may be related to reviewer explicit ratings in interesting and different ways.

This current study explores how explicit review ratings provided by the reviewer and ratings provided by the review readers are related to specific categories of words used. We focus primarily on how certain psychological categories may be connected to a review's ratings. Understanding similarities and differences in word use and review ratings may provide valuable insight into future business success and, as we detail below, cognitive processing by language users in a natural context.

II. Current Study

This study investigates how different word usage within a review text varies depending on reviewer's explicit feelings specified by a rating from 1 (negative) to 5 (positive) stars about the business reviewed. Review readers are also provided an option to rate reviews on three different dimensions—usefulness, funniness and coolness. We chose to explore how linguistic categories closely associated with psychological variables are related to explicit review ratings. To further our exploration of psychological variables that may emanate through a review text, we address how well specific categories of words found within reviews predict a reviewer's rating and the review reader's ratings of useful funny or cool. We trained a simple support vector machine (SVM) under the `e1701` library in `R` to predict review ratings based of the percentage of words within a review text that fell under certain categories.

A. Dataset

The current study analyzed approximately 229,000 reviews provided by Yelp, Inc. as part of Yelp's Dataset Challenge[3]. The Yelp, Inc. dataset consists of written reviews associated with the reviewer's explicit feelings in stars (1-5). Each review was subject to being rated as useful, funny, or cool by other reviewers among other factors. For the purpose of this study we focus on the reviewer and reader ratings, though a variety of other factors provided within the dataset such as average reviewer and business star rating, location and time of day may prove fruitful upon further inquiry.

B. Linguistic Analysis

To analyze the dataset, each review was processed using *Linguistic Inquiry and Word Count* (LIWC). LIWC scans a text

[1] Readability in Korfiatis, Garcia-Bariocanal and Sanchez-Alonso, 2012, is defined as (4.71 X (characters/words) +.5 X (words/sentences)-21.43.

[2] Hood, Hwang & King, 2013 extracted only about 100 of the top 300 keywords to represent the total number of positive words in all reviews.

[3]The dataset is provided for free at www.yelp.com/dataset_challenge

and searches for over 2,300 word stems. Words were determined to fit specific categories by independent judges over a variety of text samples from different written statements. Consistent and reliable categorical placement in over 70 different categories over time and topic lead to words chosen to represent certain categories [22]. LIWC then provides the percentage of that text that is made up of words from each of 79 categories hierarchically organized.

For example, the category *affective processes* contains the subcategory *positive emotion* consisting of 406 words including "love", "nice" and "sweet". The category of *affective processes* also includes 499 *negative emotion* words including "hurt", "ugly" and "nasty". The category *affective processes*, then, consists of 905 total words.

There are a total of 79 categories and subcategories ranging from specific uses from punctuation to cognitive words such as "think". For the purpose of this study we focus on words that fall under the category: *psychological processes.* Within this category we chose and six of its thirty-two subcategories. The words we focused on fell within one of these six categories: *social (455), positive emotion (406), negative emotion (499), cognitive (730), perception (273) and biological (567)* words. LIWC provides a considerably high dimensional understanding of word use allowing for a finer-grained analysis of how the content of a review is directly connected to a reviewer's explicit ratings.

229,206 Yelp reviews were imported and processed in Python using `json`, which was then sent to LIWC (which uses a GUI) to generate category percentages. From there, we aligned the LIWC computations with the original json structures to correlate word usage with review rating and other features. We used `nltk` and `numpy/scipy` libraries to carry out all calculations in Python, and R to build statistical models.

C. Results

We analyzed how related a reviewer star rating and a reader's review ratings were to the percentage of words within a review that fell under specific psychological categories. Each review is rated by k users as useful, funny, or cool in the theoretical range $k \in (0,\infty)$. Put differently, readers are not provided a 1-5 scale to rate reviews in these dimensions, but only indicate whether the review was, say, useful or not. As a simple exploratory analysis, we aimed to determine if a review was considered useful, funny or cool *for any reader.* Therefore, continuous variables were recoded into discrete categorical variables (e.g., reviews that were considered useful by one or more reader were recoded as useful while remaining reviews were recorded a not useful).

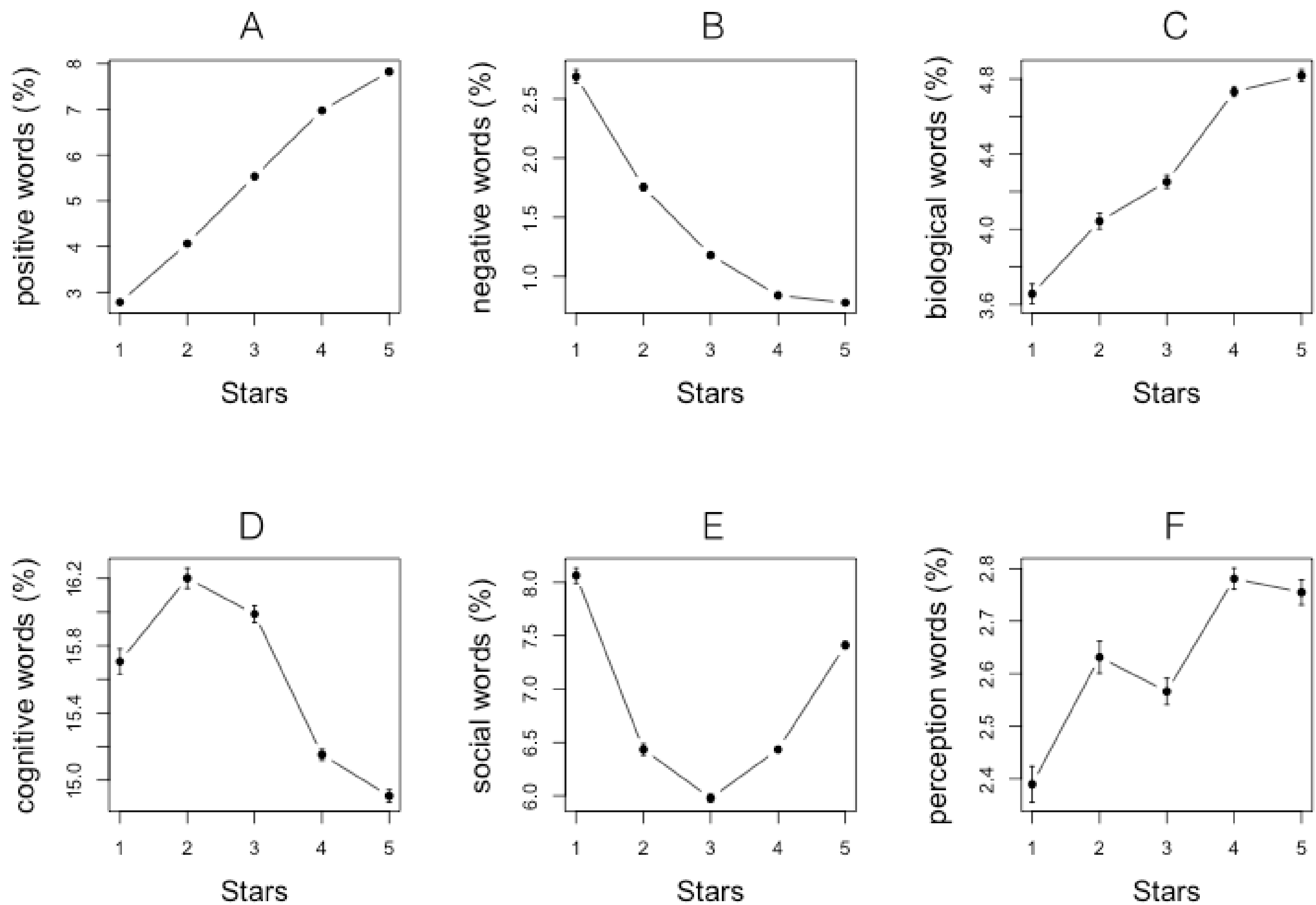

Fig. 1. All analyzed LIWC variables by star rating.

TABLE I. RATINGS BY PSYCHOLOGICAL REVIEW CONTENT

Explicit Rating by Review Content	Psychological Content					
	Positive	***Negative***	***Biological***	***Cognitive***	***Social***	***Perceptual***
Stars	$r^2 = .082$***	$r^2 = .068$***	$r^2 = .0080$***	$r^2 = .0065$***	$r^2 = .00043$***	$r^2 = .0012$***
Useful	$r^2 = .018$***	$r^2 = .00052$***	$r^2 = .0035$***	$r^2 = .00088$***	$r^2 = .000055$***	$r^2 = .000027$*
Funny	$r^2 = .019$***	$r^2 = .0041$***	$r^2 = .0041$***	$r^2 = .00024$***	$r^2 = .00015$***	$r^2 = .000085$***
Cool	$r^2 = .012$***	$r^2 = .000010$*	$r^2 = .0034$***	$r^2 = .00027$***	$r^2 = .000046$***	$r^2 = .00024$***

Table 1: R^2 And Significance (<.05, ** <.01, *** <.001) Values For Nine Liwc Variables By Yelp Factors; Star Rating, Useful, Funny And Cool. For All T-Tests Df = 229,204. Note: The Full List Of Liwc Subcategories Can Be Found At Www.Liwc.Net/Descriptiontable1.Php*

This was done so as to reflect the reader's actual discrete choice in rating reviews, but also resulted in a relatively balanced dataset (e.g., about even representation of "not cool" and "cool"). To be sure, many reviews that do not fall under one of these categories might very well be useful, funny or cool, but perhaps went unnoticed.

Table I lists the variance (r^2) within each psychological category that can be accounted for by star rating along with useful, funny and cool ratings. Star rating can account for about 15% of the variance in positive and negative emotion word occurrence, taken together. As expected, a positive relationship exists between the percentage of a review made up of positive emotion words and star rating (Fig.1A). Similarly, there was a negative relationship between negative emotion words and star rating (Fig. 1B). Of the psychological variables we chose to address, other than affective processes, collectively approximately 2% of the variance was accounted for by star rating. Of this, about 1.5% was within the categories of biological words, containing words such as "pizza", "beer" and "stomach" and cognitive words, containing words such as "think", "know" and "cause". Interestingly, the percentage of biological words within a review was positively related to star rating (Fig. 1C) while the percentage of cognitive words shows what appears to be a nonlinear relationship closely related to an inverse quadratic function (Fig. 1D). Importantly, each psychological process appears to have a unique relationship with star rating, some showing nonlinear trends such as the percentage of social words within a review (Fig. 1E). Perceptual words appear to a have a significant positive relationship with star rating (Fig. 1F) but only a very small amount of variance can be accounted for by explicit reviews ratings.

For all psychological categories, there was a significant amount of variance accounted for by usefulness ratings (Table 1). Specifically 1.8% of the variance in positive words was accounted for by usefulness ratings. Interestingly, there was a negative relationship with usefulness ratings and positive words such that readers found reviews that do not contain positive emotion words to be more useful, though they do not exactly preference negative emotion words (r^2 < .001).

TABLE II SVM MODEL ACCURACY

Explicit Ratings by Implicit Categories	LIWC Categories							
	Positive	***Negative***	***Biological***	***Cognitive***	***Social***	***Perceptual***	***Psychological Categories***	***Full Model***
Stars	*38%* *37%*	*36%* *36%*	*37%* *36%*	*35%* *35%*	*38%* *38%*	*36%* *36%*	*53%* *44%*	*72%* *49%*
Useful	*60%* *60%*	*58%* *59%*	*59%* *60%*	*60%* *60%*	*60%* *60%*	*61%* *60%*	*70%* *62%*	*72%* *69%*
Funny	*69%* *70%*	*70%* *70%*	*70%* *70%*	*70%* *70%*	*70%* *70%*	*71%* *70%*	*72%* *71%*	*75%* *72%*
Cool	*62%* *63%*	*63%* *63%*	*63%* *63%*	*63%* *62%*	*62%* *63%*	*63%* *63%*	*66%* *63%*	*73%* *68%*

TABLE II. SVM models trained on 10,000 reviews and tested 100 times on 1000 reviews. Results are presented for both training (top %) and means for 100 tests (bottom %) for each category.

Additionally, biological words account for a significant, though small, amount of variance (0.4%) such that biological words were also negatively related to useful ratings. Less than 0.1% of variance in other psychological measures were accounted for by usefulness ratings.

Funniness ratings accounted for a significant amount of variance for all variables with the most variance in positive emotion (1.9%). Similar to useful ratings, positive emotion was negatively related to funniness ratings. Specific to funniness ratings, a significant amount of variance in negative reviews (0.4%) was accounted for by funniness ratings, showing a positive relationship.

Coolness ratings accounted for a small, but significant amount of variance for all psychological categories. Importantly, and trending similarly to usefulness ratings, reviews seem to be less cool when they contain positive emotion (1.2%) and biological words (0.3%).

D. Summary

Star ratings account for the largest amount of variance in positive (8.2%) and negative (6.8%) words within reviews. But it seems obvious that one would use more positive words and less negative words when giving a higher rating and the opposite when giving a low rating. Less obvious however, psychological variables are related to star ratings in different ways. For example, social words show a quadric relationship with explicit star ratings suggesting that words like "family" and "friend" occur more frequently in reviews with extreme ratings. One could image a review that says "I would love to take my family here" or "I would never take my family here" in comparison to a more mediocre review, where sharing one's experience with family becomes less relevant. Though these effects are small, the large size of the data set allows us to detect them. These curious effects suggest that subtle aspects of language are reflecting aggregate patterns of experience by the reviewers.

In line with previous research [2] on the helpfulness of reviews, useful, funny and cool reviews show a negative relationship with positive words while significant, but small amounts of variance in other psychological categories were accounted for by these ratings as well. Additionally, biological words followed a similar negative trend. A follow up correlation revealed a significant positive relationship between positive emotion words and biological words (r = .296, p < .001). This suggests biological and positive words appear more often within similarly rated reviews. Given that a significant number of restaurants are the topic of Yelp WOM reviews, positive reviews may focus on biological processes such as how great the "food" "tastes". In this context, positive terms regarding the "great" "food" may be less useful to readers looking for more specific information such as the type of food served by the restaurant.

Words that fall under certain psychological categories are related in curious ways to specific review ratings. Beyond commercial concerns, such patterns shed light on the relationship between language and experience. In particular, social terms suggest that fundamental aspects of social experience may be encoded in the language of extreme reviews. This relates to the domain of social cognition, and recent arguments that human experience is suffused in the "social mode" [24], and exploring how this plays out in the commercial context may be theoretically interesting [25]. Results are consistent with these ideas that collective experiences may "echo" into review content.

The following section tests how well each category can predict explicit ratings. In a sense, this tests if explicit customer satisfaction can be determined indirectly through a customer's word of mouth communications.

III. SVM MODEL OF WOM

Using an SVM we tested the predictive power of less obvious linguistic aspects that make up a review. SVMs can help classify variables that may not be linearly separable; the complex array of patterns seen in Figure 1 suggests that this classifier may be suitable for this exploratory model. This is accomplished by adding more dimensions where the classification is solved. In cases where there are multiple classes such as in the case of star ratings (5 levels) a voting mechanism is used to classify cases. We used the statistical program `R` and library `e1071` to classify reviewer star rating along with the reader useful, funny and cool ratings. This particular `R` package uses a one-to-one voting mechanism rather than a one-to-all voting mechanisms [23]. The purpose for using a simple SVM was primarily in its ease of use to obtain a prediction of review rating based on certain percentages text associated with specific categories. We do not choose SVM for anything other than practical reasons, and other classifiers may suit this domain as well. For simplicity and space, we focus on these SVM results here.

Focusing on the individual psychological categories above, we predict reviewer star ratings by the percentage of words within a review that were attributed to one of these categories. In addition, we tested the predictability of all psychological categories and finally all 79 linguistic categories provided by LIWC. Table II shows the predictiveness of each psychological category for both training and test phases. The SVM model was first trained on 10,000 reviews (5% of the total dataset) then tested on one hundred sets of 1,000 reviews (1% of the total dataset). The average percent of correct classifications over the one hundred tests are reported in table II. When testing a linguistic category, the percentage associated with the ratio of words within a review, along with its associated star rating were fed into the SVM. The SVM in this sense is "trained" to recognize or dissociate a review's ratings based solely on the LIWC semantic percentages. After training, it is tested by importing novel review percentages and asked to predict what rating the review received.

A. Results

The percentage of reviews accurately categorized are presented in Table II. Approximately 37% of reviews were accurately classified by the percentage of positive words within each review. Importantly, 34% of all training reviews were given a 4 star rating. Positive emotion predicted 16% above chance and merely 3% above a model that loads onto the most frequently occurring rating. Additionally, negative words lead to an accurate classification of 36% of the data. Notably, the percentage of social words in a review lead to the accurate classification of roughly the same number of reviews as did positive emotion words. The set of psychological categories together lead to an 7% increase above positive words in predicting star rating. The full set of variables, including punctuation, personal pronouns and a variety of other linguistic categories lead to an increase in nearly 12% from positive words alone.

A relatively large percentage of reviews were correctly classified as useful, funny or cool. One possibility would be that there was significantly more reviews not rated as useful funny or cool compared to those that were. If so, the model would perform best if it were to always predict reviews based purely on frequency of review classification. For useful reviews, the ratio of useful to not useful reviews was 0.58. Positive words accurately predicted the usefulness of a review by only 2% more than classification by frequency alone. Even when controlling for this via weighting classes according to frequency the same results are found. Other psychological categories predicted roughly the same amount however, when considering all psychological categories together, prediction was 6% greater than classification by frequency. With the inclusion of all LIWC variables model prediction increased 13% above classification by frequency.

The ratio of reviews rated as funny compared to reviews that were not rated as funny was 0.30. In this case, if the model were to predict by frequency alone, we would anticipate the model's classification accuracy to be 70%; classifying all cases as not funny. Our results reflect this finding, with little to no change with the addition of psychological variables and only a very slight increase with the full model. This finding remained after weighting cases.

The ratio of "cool" to "not cool" reviews was 0.37. Given this, if the model were to always predict that reviews would not be "cool" then the percentage of reviews predicted as cool would be roughly 63%. As shown in Table 2, this is precisely what was predicted. Inquiring into the predictions of the model, nearly all reviews were in fact predicted as not cool. The inclusion of all psychological variables did not increase the predictability of classifying reviews as cool or not cool, though a slight increase (5%) did occur with the inclusion of all LIWC categories. This finding remained even after balancing cases.

B. Summary

The SVM model accurately predicted reviewer star ratings for roughly one third of the test data; this was 15% above chance, and 3% above frequency from positive reviews alone. While other psychological categories did not increase the predictability of review classification independently, together 7% more reviews were accurately classified during testing. This increase suggests different psychological categories account for some amount of variance in review ratings beyond that accounted for by positive words within reviews. An additional 5% classification accuracy was obtained using the full set of LIWC categories. Though not presented individually here, but grouped in with the full model, other word groups such as personal pronouns [16] may provide additional predictive power in explicit rating classification tasks. A more thorough analysis, outside the scope of the current study, may be required to further understand what might be contributing to model accuracy

When predicting useful reviews, positive emotion predicts only 2% above classification by frequency. Other psychological variables did not provided any more accurate classification above frequency. However, when considering all LIWC variables there was a slight increase in accuracy. Similar trends were seen for classifying cool and funny reviews.

IV. General Discussion

Explicit business review ratings provide insight into reviewer customer satisfaction, while review reader ratings measure how successful review cites are at communicating desired information to its customers. In both cases, interesting patterns appear when considering the underlying psychological processes nested within a review. Understanding how such processes can help to elucidate what customer values influence customer satisfaction [4].

In this exploratory analysis we used a tools from Natural Language Processing to uncover potentially interesting behavioral and emotional content nested within WOM reviews. We show that a reviewer's star rating along with reader ratings of useful, funny and cool account for a significant amount of variance in positive words explicitly. Though this may seem intuitive for star ratings, it is not necessarily obvious that positive words were negatively related to usefulness ratings, though only slightly. This adheres to previous research where reviews landing somewhere between extremely positive and extremely negative were considered most helpful [2]. Relatedly, a novel finding revealed biological words were positively correlated with positive words, and both maintain a negative relationship with usefulness.

Even less intuitive: Explicit review ratings account for a significant amount of variance in psychological words that make up a review text. Previous work shows variance in psychological word usage is highly related to underlying psychological processes such as depression [16], longevity [20] and social structure [17,18]. While this connection is not explicit, we speculate that understanding the language reviewer's use acts as a window into the reviewers' psychological processes underlying her explicit review ratings. In particular, as we discussed above, the relationship of social terms to review rating is predicted by theories that human cognition is fundamentally social. Curiously, reviewers who have a more intense experience, whether positive or negative, are significantly more likely to mention others in their WOM evaluation. These findings in general suggest that future analysis of large-scale data sets of this kind could further corroborate theories in cognitive science while also being useful in the commercial context [25].

When considering how well psychological categories predict review ratings, both positive and negative word perform roughly the same as other psychological categories; though this was not much better than chance. When considering the impact of all psychological categories there is an increase in star rating predictability. Interestingly no single category provides a significant change in review rating predictability while collectively the amount of variance accounted for by all psychological categories provides substantially more predictive power. The subtle amounts of variance in psychological processes account for a substantial amount of predictive power in review ratings. Access to such processes via word usage in WOM reviews can provide valuable insight toward establishing better customer-business relations.

Explicit reviewer ratings provide an obvious mark of customer satisfaction. Touted as the "missing link" between customer satisfaction and business success [3] WOM reviews, when paired with explicit review ratings, take on a new level of complexity. A variety of studies have focused on uncovering on how WOM reviews relate to explicit ratings [1,2,4,5,6,9,10,11], yet data remain mixed. No single analysis has proven to convincingly report on the underlying structure behind WOM reviews as they relate to customer satisfaction. Provided this assumption, we explored and analyzed a variety of categories closely related to psychological processes such as cognitive and social words.

Our results suggest, rather naturally, that different categories of words hold different relationships with explicit reviewer ratings. Interestingly some unknown relationships were discovered; such as the correlation between positive and biological words in WOM reviews, and how they related to usefulness ratings. We speculate that understanding how WOM reviews are related to explicit review ratings requires a complex and comprehensive approach; one that adheres to a variety of different methodologies. Indeed, a multitude of previous research using a variety of methods have shown interesting connections between WOM reviews and explicit review ratings including review readability [11], complexity [10], frequency of occurrence [8] and now word use. The current exploratory analysis shows how somewhat disparate subtleties within review text, overlooked by explicit review ratings, can provide insight into specific reviewer values underlying customer satisfaction.

REFERENCES

[1] A. Serra Cantallops and F. Salvi, "New consumer behavior: A review of research on eWOM and hotels" International Journal of Hospitality Management, Vol. 36, pp. 41-51, 2014.

[2] S. M. Mudambi and D. Schuff. "What makes a helpful online review? A study of customer reviews on Amazon. com." MIS quarterly, vol. 34, no.1, pp. 185-200, 2010.

[3] F. F. Reichheld, "The one number you need to grow." Harvard business review, Vol. 81, no. 12, pp. 46-55, 2003.

[4] A. B. Eisingerich, S. Auh and O. Merlo, "Acta Non Verba? The Role of Customer Participation and Word of Mouth in the Relationship Between Service Firms' Customer Satisfaction and Sales Performance." Journal of Service Research, Vol. 17, No. 1, pp. 40-53, 2014.

[5] Q. Ye, R. Law and B. Gu, "The impact of online user reviews on hotel room sales." IJHM, Vol. 28, pp. 180–182, 2009.

[6] [S. H. Jun, C. A. Vogt, K. J. MacKay, "Online information search strategies: a focuson flights and accommodation." J. Travel Tourism Mark, Vol. 27, pp. 579–595, 2010.

[7] M. R. Bowers, C. L. Martin and A. Luker, "Trading places: employees as customers, customers as employees." Journal of Services Marketing, Vol. 4, No. 2, pp. 55-69, 1990.

[8] E.W. Anderson, "Customer satisfaction and word of mouth." Journal of service research, Vol. 1, No. 1, 5-17, 1998.

[9] M. Luca, "Reviews, reputation, and revenue: The case of Yelp. com," Harvard Business School, 2011.

[10] D. W. Vinson and R. Dale, "Valence weakly constrains the information density of messages." In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), Proceedings of the 36th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society. 2014.

[11] N. Korfiatis, E. García-Bariocanal, and S. Sánchez-Alonso, "Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content," Electron. Commer. Res. Appl., vol. 11, no. 3, pp. 205–217, 2012.

[12] S. Feng, R. Banerjee, and Y. Choi, "Syntactic stylometry for deception detection," in Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2, pp. 171–175, 2012.

[13] B. Hood, V. Hwang, and J. King, "Inferring Future Business Attention." retrieved from http://www.yelp.com/dataset_challenge

[14] J. W. Pennebaker, C. K. Chung, M. Ireland, A. Gonzales, and R. J. Booth, "The development and psychometric properties of LIWC2007," Austin TX LIWC Net, 2007.

[15] J. W. Pennebaker, T. J. Mayne and M. E. Francis, "Linguistic predictors of adaptive bereavement." J. Personal. Soc. Psychol. Vol. 72, pp. 863–71, 1997.

[16] W. Bucci and N. Freedman, "The language of depression." Bull. Menninger Clin.Vol. 45, pp. 334–58, 1981.

[17] E. Goffman, "The Presentation of Self in Everyday Life." Garden City, NY: Doubleday. 1959.

[18] G. Lupyan and R. Dale, "Language structure is partly determined by social structure" PLoS ONE, Vol. 5, e8559, 2010. doi:10.1371/journal.pone.0008559

[19] J.W. Pennebaker and M.E. Francis, "Cognitive, emotional, and language processes in disclosure." Cogn. Emot. Vol. 10, pp. 601–26, 1996.

[20] D.D. Danner, D. A. Snowdon W. V. Friesen, "Positive emotions in early life and longevity: findings from the Nun Study." J. Personal. Soc. Psychol. Vol. 80, pp. 804–13, 2001.

[21] M. Hu and B. Liu, "Mining and summarizing customer reviews," in Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 168–177, 2004.

[22] J.W. Pennebaker and L.A. King, "Linguistic styles: language use as an individual difference." J. Personal. Soc. Psychol. Vol. 77, pp. 1296–312, 1999.

[23] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," Neural Networks IEEE Trans., vol. 13, no. 2, pp. 415–425, 2002.

[24] M. Gallotti and C. D. Frith, "Social cognition in the we-mode." *Trends in cognitive sciences*, Vol. 17, No. 4, pp. 160-165, 2013.

[25] P. M. Hart and R. Dale, "With or without you: The positive and negative influence of retail companions." Under review.