Understanding the Yelp Review Filter: An Exploratory Study

David Kamerer
Loyola University Chicago, dkamerer@luc.edu

Recommended Citation
Understanding the Yelp review filter: An exploratory study
by David Kamerer

Abstract
Reviews on Yelp.com can be an important factor in driving customers to a business. However, many business owners have expressed concern with Yelp's review filtering system, which was created to flag low-quality or fake reviews. This study performs a content analysis of a subset of Yelp restaurant and religious organization reviews, visible and filtered, exploring signals from the reviews or the reviewers that might explain the filtering process. The study finds that factors intrinsic to the review itself are not related to filtering, but factors related to the reviewer are strong predictors. The Yelp system is much more likely to filter reviews from occasional, isolated reviewers than from prolific, socially connected reviewers.

Contents
Introduction
Economic theory applied to reviews
About Yelp
Complaints about Yelp
Yelp's position
Research questions
Method
Results
Discussion and conclusion

Introduction
"Please help the business people from the Internet Mafia," said a business owner, complaining about Yelp.com to the U.S. Federal Trade Commission (FTC). "Yelp is the thug of the internet," wrote another. Yet another business owner complained that 42 out of 47 reviews of its business abruptly vanished (Kang, 2013). These were among almost 700 complaints against Yelp filed at the FTC.

Business owners know: online reviews are powerful. Research supports this view as well. Positive reviews on Yelp have been empirically shown to drive business success. One study found that a one-star increase in a restaurant's Yelp rating brought about a five- to nine percent increase in revenue (Luca, 2011). Another study found a one-half star increase in a restaurant's Yelp rating was linked to selling out 49 percent more frequently (Anderson and Magruder, 2012).

It's not just bad reviews that upset business owners, however. Almost all reviews posted to Yelp are viewable online, but some — those that Yelp's software views as "not recommended" — are hidden from most viewers and are not counted towards the star rating. When a business finds a positive review that has been filtered, the owner is certain to take exception.

Yelp has steadfastly maintained that the review filter is an algorithm designed to keep fake reviews from hurting the integrity of its database. The company further says that no individual has the power to move reviews into or out of the filter. But Yelp's transparency — the company allows anyone to see the filtered reviews — has undermined trust. Business owners cry foul when positive reviews are filtered, and also when negative reviews are not. But there's also a more serious accusation: that Yelp is a pay-for-play scam, and that the company manipulates the review filter to reward people who buy advertising packages and to punish those who don't.

This view is widely held in the small business community, and has grown with class-action lawsuits against Yelp, media coverage from aggrieved business owners, and complaints against Yelp filed at the Federal Trade Commission.

This study examines the Yelp review filter to better understand how it works. The inquiry is prompted by the chasm between Yelp's stated position and the many vocal complaints from businesses about Yelp.

Economic theory applied to reviews
Online reviews are best understood through economic models. They are part of the Web 2.0, which is characterized by sites where consumers can publish content on socially-connected platforms. While many aspects of social publishing may be appealing to consumers, one key characteristic of this activity relevant to publishers is that it creates free content. This model, where status updates, tweets and blog posts create the content that builds the inventory of viewers, which is then monetized by ads, has fueled the growth of Facebook, Twitter, Huffington Post and many other sites. In this case, Yelp economically benefits from crowd-sourced reviews because the company does not have to pay its own staff to write reviews. Instead, it has created a community with social incentives, such as "friend" and review endorsements (these include "useful," "funny" and "cool"). Yelp's most active reviewers are granted "elite" status on the site. Members of the Yelp Elite Squad are invited to private events and receive rewards from Yelp to reinforce their behavior.

Consumers turn to reviews in part to reduce information asymmetry. This concept has been developed by economists to challenge the assumption of a rational market (Akerlof, 1970). Typically, sellers are information-rich while buyers lack information. Thus, "buyer behavior is crucially dependent upon the information that is available before and after purchase. In order to make choices, buyers need to at least know the price and quality of the various alternatives that they are considering" (Nayyar, 1990).

Lacking information about the best choice of business to patronize, consumers use many signals to reduce information asymmetry. Traditionally, this has included consulting owned media (such as a Web site or brochure) and paid media (advertising). For example, a restaurant may use words like "organic," "free-range" or "local" to communicate the type and quality of food offered to diners.

Digital ads, however, have limited effectiveness, particularly in getting people to respond to a call to action. While they may work as a branding tactic, industry-wide, banners receive a less than one-tenth of one percent click rate (Chapman, 2011). In a study conducted by Edelman for Adobe, 53 percent of respondents reported that they think "most marketing is a bunch of BS," 68 percent said that online advertising was annoying, and 54 percent said they didn't think that online banners worked (Adobe, 2012).

As social media have become mass media channels, peer-based or crowd-sourced messages such as online reviews have become an important force in shaping opinions. Edelman's (2012) annual Trust Barometer research has consistently shown that we're most likely to trust "a person like me." Crowd-sourced reviews have surfaced across product categories, including consumer products (Amazon), movies (Rotten Tomatoes) and hotels (Trip Advisor). One study found in a quasi-experimental comparison of identical books on Amazon and Barnes and Noble Web sites that more positive reviews for a book on one site were related to higher sales of that book on that site (Chevalier and Mayzlin, 2006). Reviews reduce information asymmetry between buyer and seller. As such, in cases where there is relatively little asymmetry, consumers value reviews less. For example, in one study, positive reviews were helpful for independent restaurants, but not for chain restaurants. This study also found that chain restaurants have lost market share as the online review ecosystem has grown in importance (Luca, 2011). Both of these findings speak to the value of reviews in bridging a competitive gap between chains, which may be familiar to consumers, and independent restaurants, which are likely not.
Filtered reviews on Yelp were found to be more extreme in rating (more likely to be one-star or five-star) than visible reviews (Luca and Zervas, 2013). The study also found that restaurants are more likely to commit review fraud when their reputations are weak (i.e., they have few reviews or poor reviews), and that restaurants are more likely to leave fake reviews for competition when they are in a highly competitive marketplace.

Another economic concept related to online reviews is signal cost. Sometimes the cost is low, such as when anyone can write a review (this is true on Amazon and Yelp, for example). However, a review posted on Open Table, where you must have made reservations in order to be eligible to write a review, has a higher signal cost. Similarly, Expedia Travel has a "paid and stayed" requirement for reviewers, imposing a higher signal cost. On Angie's List, where you must pay for a subscription, the signal cost for all reviews is rolled into an annual membership fee. Sites with lower signal costs have been found to be more likely to host corrupt or fake reviews (Ott, et al., 2012). Sites with higher signal costs reduce the overall number of reviews written, trading the lower volume for higher levels of trust.

Another economic concept applied to reviews is the exposure benefit. All things equal, a site with more viewers has a higher exposure benefit. Thus, sites with low signal costs and high exposure benefits are most likely to have corrupt reviews. In a study of six different travel review sites, the authors observed that a higher signal cost did reduce the number of corrupt reviews (Ott, et al., 2012).

Yelp reviews feature low signal cost (anyone can create a free account and write a review) with high exposure benefit (Yelp has the largest audience of all review sites). Based on these two factors, there is likely to be a high proportion of corrupt reviews on Yelp, relative to other sites.

### About Yelp

Yelp is the largest standalone review site online, home of more than 40 million reviews and garnering more than 100 million unique visitors per month (Stoppelman, 2013).

A business may be completely passive in managing its presence on Yelp, or may take a more active role. A third option is to also advertise on Yelp.

At the simplest level, the business can learn from the reviews to improve its performance.

The business owner can also claim its business. This allows the owner to annotate the page by adding photos, information on location, hours of operation and other relevant details. The business owner also gains access to analytics to track views on the site. Additionally, the owner can respond to reviews or create Yelp Deals or a mobile check-in offer to encourage customer visits.

The business can also advertise on Yelp. Fundamentally, Yelp is an advertising company and it monetizes its operations by selling its inventory of reviewers and readers to advertisers. Advertising packages start at US$300/month, and are based upon a guaranteed number of impressions. Advertisers can upload show slides and a video to their Yelp page and can add a "call to action" button (such as "reserve now"). Ads are served on relevant searches by location and category, appearing in competitive searches. Yelp also removes competitor ads removed from the pages of advertisers.

The ad programs deliver poor value, according to online marketer Rakesh Agrawal (2012). Yelp's advertising charges for impressions rather than clicks. Across the Internet, impressions vary in cost but average about US$2.52 per 1000 impressions, or CPM (Lipsman, 2010). Because of the abundance of online ad outlets, CPMs are often well below US$1. According to Agrawal (2012), "At a time when much online advertising is being sold for 60 cents per thousand impressions (CPMs), Yelp is charging some local advertisers $600 per 1,000 impressions. At the high end, it's a $600 CPM. At the low end, that's a still eye-popping $367 CPM — more than 10–times the rate of a Super Bowl ad." Yelp ad packages also require a three, six or 12-month contract, forcing the business to continue the program even if it is not effective.

### Complaints about Yelp

Many business owners resent Yelp as an advertising platform. They may not like to pay for ads when they can't control how people rate their business or what ads may appear on their page. And they may not like dealing with pushy ad salespeople.

But the strongest emotions are reserved for Yelp's review filter. The filter controls which ads are easily viewed and counted towards the overall star rating. Unlike other review sites, Yelp makes it possible to see "suspect" reviews (to view filtered reviews, scroll to the bottom of a 'business' page). Click on the light gray link that says "... reviews that are not currently recommended"). When a business owner sees a positive review that has been filtered, he or she wants to have it made visible. And the business owner may also wish to hide negative reviews in the filter.

Many small business owners claim that Yelp uses the review filter to reward advertisers and to punish everyone else. Many of these complaints allege extortion. They claim that a Yelp salesperson will make promises of moving positive reviews out of the filter when a business commits to a Yelp advertising plan. Alternately, say the complainers, the salesperson may threaten to move negative reviews out of the filter for a business that declines the plans.

These are dominant themes in complaints filed against Yelp at the Federal Trade Commission. These complaints were obtained in a Freedom of Information request by MuckRock.com (2012) (https://www.muckrock.com/), an investigative Web site. Forbes (2013) summarizes some of these complaints [1]:

> "Consumer states that [Yelp] told him to pay 300.00 for a 'membership' that would give him more positive reviews. Consumer noticed that a couple of positive reviews popped up immediately after paying the money."

> "Although the site cites that disclosing those [filtering] criteria will allow circumventing their built-in checks and balances, in practice it gives them a complete power to manipulate the ratings of any business."

> "The site constantly and relentlessly filters out the good and excellent reviews displaying less positive reviews causing the rating to downgrade."

Over four years, almost 700 businesses complained about Yelp to the FTC. In addition to the extortion charges, there are complaints about the cost of advertising packages, high-pressure sales people, listing of home addresses and phone numbers in the database, out-of-date listings, poor customer service and refusal to remove reviews, even when the facts were proven to be false.

These complaints have also been the subject of journalistic investigations. In an article entitled "Yelp and the business of extortion 2.0," The East Bay (California) Express interviewed six business owners who claimed that their "Yelp sales representatives promised to move or remove negative reviews if their business would advertise. In another six instances, positive reviews disappeared or negative ones appeared — after owners declined to advertise." The article also noted "And in at least one documented instance, a business owner who refused to advertise subsequently received a negative review from a Yelp employee" (Richards, 2010).

The claim of extortion has also been taken up by other journalists and has even been featured as a storyline on the television show "The People's Court" (Jones, 2010; Chang, 2011). This is in part due to the expansive groups such as "I hate Yelp" (Yelp–Sucks, 2014; BOSSHi, 2014; I Hate Yelp, 2014).

Further evidence of anti-Yelp sentiment from businesses is manifest in Web sites, such as Yelp–sucks.com, the "Yelp survival guide," at BOSSHi.com, and Facebook groups such as "I hate Yelp" (Yelp–Sucks, 2014; BOSSHI, 2014; I Hate Yelp, 2014).

While it’s easy to see a trust deficit surrounding Yelp’s practices, thus far Yelp has not lost a major legal challenge (Chang, 2011). This is in part due to the expansive treatment online businesses receive under Section 230 of the Communications Decency Act, which generally absolves businesses for liability from content posted by third parties (Title 47, Section 230). But Yelp’s critics have also failed to prove their cases.
Yelp's position

When a site contains 40 million reviews, a few fake reviews are bound to get through. These might include a merchant writing a positive review of his or her own business, a merchant writing a negative review of a competitor’s business, or a disgruntled employee/ex-spouse/troll writing a negative review of almost anything.

Yelp is also threatened by a cottage industry of reviewers—for–hire, doing business online at places like Craigslist, Fiverr.com and through the Amazon Mechanical Turk marketplace.

Yelp has a duty to protect the integrity of its database, as do other review sites. But Yelp has drawn a disproportionate amount of criticism for its approach to filtering reviews. By comparison, Amazon, another large review database, reserves the right to remove reviews in four different categories: objectionable material, promotional content, inappropriate content and off–topic information (Amazon, 2014). You could argue that Yelp’s practice is more transparent because if you want, you can read the suspect reviews and decide for yourself.

Yelp relies on algorithm–derived software to sort the reviews into live or filtered. The company has repeatedly stated that it does not manually filter reviews. On a Reddit AMA ("ask me anything") CEO Jeremy Stoppelman said, “There has never been any amount of money you could pay us to manipulate reviews. We do have an algorithm that highlights the most useful and reliable reviews on our site which is about 75 percent of contributed content” (Wasserman, 2013). Wrote Yelp spokesperson Vince Sollitto: “Some news outlets have recently run stories or columns re–hashing the sensational allegation that Yelp manipulates reviews and ratings to reward advertisers or punish non–advertisers. Let me be clear: This claim is not — and has never been — true” (Fell, 2013).

Yelp’s explanation of filtering is shown to consumers every time they request a review that’s not recommended (filtered). To see this video, scroll to the bottom of the first page for any business that’s on Yelp. When you click on the grey "reviews that are not currently recommended" the video is on the top of the next page served. In part, the voiceover says:

"Every Yelp review is automatically evaluated by Yelp's recommendation software, based up quality, reliability and user activity on Yelp. More often than not, those useful reviews come from active members of the Yelp community. Our recommendation software is looking for reviews written by people whose opinions, positive or negative, we feel are helpful and reliable to consumers.

Currently, about 75 percent of all reviews are recommended. We try not to highlight reviews written by users we don’t know much about, or those that might be fakes, or unhelpful rants or raves” (YouTube, 2013).

Research questions

This study investigates how the Yelp review filter functions in a small subset of reviews on the site. The study analyzes how the filter functions at the reviewer level. The larger question is: does Yelp filter the business or the reviewer? Most complaints against Yelp assume that Yelp filters the business. For example, a critic may believe that a business that declined an advertising package might see negative reviews moved from the filter to the live database of reviews. The first question, does Yelp filter the business, is difficult to answer because it requires information about whether the business is a Yelp advertiser, and then time–based analysis based upon when the business began advertising or declined an advertising offer. But it’s much simpler to measure whether Yelp filters the reviewer. Essentially, this is Yelp’s position. Of the two questions, it’s more easily tested. Many of the signals about the reviewer are posted on the Yelp Web site. Some, such as the reviewer’s IP address, are not. But there is plenty of information about reviewer activity and whether reviews are filtered, so we will begin there.

We will investigate two research questions:

1. Are the review–based signals related to the review being filtered?
2. Are the reviewer–based signals related to having the review being filtered?

Method

Data for this study were collected via content analysis of reviews written in two categories: restaurants located on Chicago’s north side and religious organizations across Chicago. The grouping of restaurants was chosen for having a low concentration of chain restaurants; previous research has shown that reviews are less important for chain restaurants because consumers often know what to expect before dining at one. The religious organizations category was chosen for contrast; additionally, the review volume in this category is much lower than it is for restaurants.

The restaurants sample was selected in January 2014, and was created in two stages. First, researchers established a sampling frame of 809 restaurants in the Yelp database that met the criteria of being restaurants on Chicago’s north side. The sample was created using a skip interval of five and a random start point created by the Web site random.org ([http://www.random.org/]), yielding 162 eligible restaurants. Once the restaurants were selected, the total number of reviews for each restaurant was tabulated. This number was created by adding all visible reviews to all filtered reviews. Then one review was selected from this restaurant using a number generated by the random.org site.

This scheme was adopted as a relatively random way to gain access to the database of reviews. It should be noted that, since one review was pulled from each restaurant in the sample, that heavily reviewed or popular restaurants are under–represented in the study. Because we’re not interested in reviews at the restaurant level (and instead interested in the reviewer), this is only a minor sampling limitation.

The restaurants sample was selected in January 2014, and was created in two stages. First, researchers established a sampling frame of 809 restaurants in the Yelp database that met the criteria of being restaurants on Chicago’s north side. The sample was created using a skip interval of five and a random start point created by the Web site random.org ([http://www.random.org/]), yielding 162 eligible restaurants. Once the restaurants were selected, the total number of reviews for each restaurant was tabulated. This number was created by adding all visible reviews to all filtered reviews. Then one review was selected from this restaurant using a number generated by the random.org site.

The restaurants sample was selected in January 2014, and was created in two stages. First, researchers established a sampling frame of 809 restaurants in the Yelp database that met the criteria of being restaurants on Chicago’s north side. The sample was created using a skip interval of five and a random start point created by the Web site random.org ([http://www.random.org/]), yielding 162 eligible restaurants. Once the restaurants were selected, the total number of reviews for each restaurant was tabulated. This number was created by adding all visible reviews to all filtered reviews. Then one review was selected from this restaurant using a number generated by the random.org site.

This scheme was adopted as a relatively random way to gain access to the database of reviews. It should be noted that, since one review was pulled from each restaurant in the sample, that heavily reviewed or popular restaurants are under–represented in the study. Because we’re not interested in reviews at the restaurant level (and instead interested in the reviewer), this is only a minor sampling limitation.

The religious organizations sample was drawn in August 2014. Because there are fewer reviews in this category, the list was sorted from most– to least–reviewed. The sample was built from the top down, using random.org to choose one review from each organization. For the religious organizations sample, n=128.

The variables related to the review in question were coded as follows:

- Word count of review (each review was pasted into Microsoft Word, which calculated the number of words)
- Number of stars of review (1–5 stars)
- Review filtered? (yes/no)

The variables related to the reviewer were coded as follows:

- Yelper name
- Number of Yelp friends
- Number of Yelp reviews by reviewer
- Avatar picture provided (yes/no)
- Number of fans of Yelper
- Number of reviews written by Yelper at each star level
- Number of “useful,” “funny,” and “cool” votes for Yelper

After the sample was created, preliminary analysis showed a low percentage of filtered reviews (nine percent). In order to increase the incidence of filtered reviews, researchers revisited the sample and added cases likely to be filtered. This was done by calculating the 93rd percentile review for each organization in the sample and coding that review. The resulting filtered reviews were then added to the database, yielding a total sample of 240 reviews. The resulting database, then, is weighted to include a disproportionate percentage of filtered reviews. While Yelp reports about 25 percent of all reviews to be filtered, in this study the percentage is 35 percent for restaurants and 38 percent for religious organizations.

The principal investigator contributed to the database and trained two additional coders. Because each coder only reported manifest content (such as writing down the number of stars of a review or the number of reviews written by an individual), no intercoder reliability check was needed. In essence, no judgments were recorded; no interpretations were made. The coders simply transferred numbers and text from the Yelp database to the study database.

These data were keypunched into Microsoft Excel and then imported to IBM SPSS 20 for analysis.
It is important to note that the Yelp database is live and constantly changing. New restaurants may be added, closed restaurants may be removed, and consumers are constantly adding new reviews. Reviews may also be moved into and out of the filter. The database presents two snapshots of the Yelp ecosystem, from one city (Chicago) and two categories (north side restaurants and religious organizations).

Results

Results are divided into two sections: review–based variables and then reviewer–based variables.

Table 1 shows the word counts in the reviews under study. Overall, visible reviews are longer. In the restaurant sample, filtered reviews averaged 120 words, while visible reviews averaged 144 words. In the religious organizations sample, filtered reviews averaged 99 words, while visible reviews averaged 150 words. Overall, filtered reviews for restaurants are 83 percent as long as visible reviews; for religious organizations, filtered reviews are 66 percent as long as visible reviews.

<table>
<thead>
<tr>
<th></th>
<th>Restaurants</th>
<th>Religious organizations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Filtered</td>
<td>Visible</td>
</tr>
<tr>
<td>Words</td>
<td>120</td>
<td>144</td>
</tr>
<tr>
<td>SD</td>
<td>139</td>
<td>128</td>
</tr>
<tr>
<td>N</td>
<td>85</td>
<td>147</td>
</tr>
</tbody>
</table>

Table 2 shows the average star ratings of the reviews in the study. Overall, reviews are generally positive. In the restaurant sample, filtered reviews averaged 3.5 stars (one star is lowest and five stars highest), while visible reviews averaged 3.6 stars. Both groups of reviews had the same standard deviation of 1.3, so the variability of reviews in both groups is about the same. Reviews are strongly positive in the religious organizations sample, with a mean of 4.5 stars for filtered reviews and 4.7 stars for visible reviews. Variability is slightly higher (SD=1.1) for filtered reviews than for visible reviews (SD=.8).

<table>
<thead>
<tr>
<th></th>
<th>Restaurants</th>
<th>Religious organizations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Filtered</td>
<td>Visible</td>
</tr>
<tr>
<td>Stars</td>
<td>3.5</td>
<td>3.6</td>
</tr>
<tr>
<td>SD</td>
<td>1.3</td>
<td>1.3</td>
</tr>
<tr>
<td>N</td>
<td>84</td>
<td>147</td>
</tr>
</tbody>
</table>

A word cloud analysis was conducted to explore review themes between categories. The online service Wordle.net (http://www.wordle.net/) was used to visualize the predominant words in filtered reviews, and then again to look at just visible reviews. Figure 1 shows the word cloud for restaurant filtered reviews; Figure 2 shows the word cloud created by restaurant visible reviews only.

Figure 1: Word cloud for all filtered restaurant reviews.
The two word clouds are remarkably similar, with neutral descriptive words, such as "food," "place," "good," "great," "like," and "service" dominating. The interior words are reflective of the different restaurants and kinds of food that were the subjects of the reviews. No "trigger words," for example, superlatives that might signal a fake review, are evident in either word cloud.

The word cloud for religious organization filtered reviews is presented in Figure 3, while the visible reviews are presented in Figure 4. As in the restaurant review word clouds, the two seem very similar, with words like "church," "people," "place," "community" and "great" dominating.
While the word clouds do offer a good overall visualization, they fail to discriminate between the filtered and visible reviews. However, it is possible to assess quantitative differences between the two groups using a quantitative approach. Linguistic Inquiry and Word Count (LIWC) software \cite{2} “calculates the degree to which people use different categories of words.” The software matches the text in question with a dictionary that has been validated across many studies. It tabulates manifest content, like counting words and sentence length, as well as more subtle analyses, such as counting swear words or “social processes.” In all, LIWC can analyze more than 99 categories of text. Every review in this study was analyzed with LIWC software and then these new results were appended to the original data file.

Would quantitative study of the actual words provide more discrimination between visible and filtered reviews? For this study, we looked at two groupings of variables. The first consists of structural variables, which include word count, words per sentence, six–letter (and longer) words, the voices “I,” “we” and “they,” and the tenses “past,” “present” and “future.”

Table 3 shows the structural variables used in restaurant reviews. While there are minor differences in the occurrences of these variables, independent sample T–tests show only variable with a significant difference. Reviews using “they” are significantly more likely to be filtered (1.4 occurrences in filtered reviews compared to .8 occurrences for visible reviews). All of the other variables showed no difference between the filtered and visible groups.

| Table 3: Structural variables and filtering of restaurant reviews. |
|-------------------|-----------------|--------|-------|--------|
| variable          | visible | filtered | t     | df    | sig    |
| Word count        | 125    | 118      | 0.422 | 221   | 0.671  |
| Words per sentence| 15     | 16       | -0.242| 221   | 0.809  |
| Six letter words  | 17     | 16       | 0.527 | 221   | 0.598  |
| I                 | 3      | 4        | -0.772| 221   | 0.441  |
| We                | 1      | 1        | -0.313| 221   | 0.755  |
| They              | 1      | 1        | -2.731| 221   | 0.007  |
| Past              | 5      | 5        | -0.918| 221   | 0.360  |
| Present           | 7      | 7        | 0.277 | 221   | 0.782  |
| Future            | 1      | 1        | 1.023 | 221   | 0.307  |

Table 4 shows the structural variables used in reviews of religious organizations. Again, using independent sample T–tests, there was no difference between the structural variables as used in visible reviews and filtered reviews.

| Table 4: Structural variables and filtering of religious organization reviews. |
|-------------------|-----------------|--------|-------|--------|
| variable          | visible | filtered | t     | df    | sig    |
| Word count        | 148    | 102      | 2.300 | 125   | 0.023  |
| Words per sentence| 17     | 17       | -0.152| 125   | 0.879  |
| Six letter words  | 20     | 19       | 1.025 | 125   | 0.307  |
| I                 | 4      | 3        | 0.621 | 125   | 0.536  |
| We                | 1      | 0        | 1.069 | 125   | 0.287  |
| They              | 1      | 1        | 0.300 | 125   | 0.765  |
| Past              | 3      | 3        | 0.965 | 125   | 0.337  |
| Present           | 8      | 8        | -0.197| 125   | 0.844  |
| Future            | 0      | 1        | -1.285| 125   | 0.201  |

A meta–analysis of studies that utilized LIWC provides insight into its text categories. Several have been linked to deception, although the construct may have been operationalized in different ways. According to the meta–analysis, “across the studies when participants were lying they used more negative emotion, more motion words (e.g., arrive, car, go), fewer exclusion words, and less first–person singular” (Tausczik and Pennebaker, 2010). The studies also show that liars use a higher word count, less first–person singular voice, and more “sense” words. “Motion, exclusion and sense words all indicate the degree to which an individual elaborated on the description of the scenario” (Tausczik and Pennebaker, 2010).

To test these general observations about lying and deception, independent sample T–tests were run on the variables negative emotion, exclusion, motion, and the three sense variables “see,” ”hear” and “feel,” with filtered/visible as the dependent variable.
Table 5 shows the content variables in the restaurant review sample. Not one variable is significantly related to filtering. This analysis is repeated with reviews of religious organizations in Table 6 with the same result; there is no relationship between the content variables and whether reviews were filtered.

<table>
<thead>
<tr>
<th>variable</th>
<th>visible</th>
<th>filtered</th>
<th>t</th>
<th>df</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative emotions</td>
<td>1.34</td>
<td>0.91</td>
<td>0.876</td>
<td>221</td>
<td>0.382</td>
</tr>
<tr>
<td>Exclusion</td>
<td>2.80</td>
<td>2.89</td>
<td>-0.251</td>
<td>221</td>
<td>0.382</td>
</tr>
<tr>
<td>Motion</td>
<td>1.76</td>
<td>1.76</td>
<td>-0.039</td>
<td>221</td>
<td>0.969</td>
</tr>
<tr>
<td>See</td>
<td>0.73</td>
<td>0.50</td>
<td>1.642</td>
<td>221</td>
<td>0.102</td>
</tr>
<tr>
<td>Hear</td>
<td>0.43</td>
<td>0.49</td>
<td>-0.418</td>
<td>221</td>
<td>0.676</td>
</tr>
<tr>
<td>Feel</td>
<td>0.60</td>
<td>0.39</td>
<td>1.603</td>
<td>221</td>
<td>0.110</td>
</tr>
</tbody>
</table>

Table 6: Content variables and filtering of religious organization reviews.

<table>
<thead>
<tr>
<th>variable</th>
<th>visible</th>
<th>filtered</th>
<th>t</th>
<th>df</th>
<th>sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative emotions</td>
<td>0.71</td>
<td>0.65</td>
<td>0.347</td>
<td>125</td>
<td>0.729</td>
</tr>
<tr>
<td>Exclusion</td>
<td>2.35</td>
<td>1.96</td>
<td>-0.251</td>
<td>221</td>
<td>0.382</td>
</tr>
<tr>
<td>Motion</td>
<td>1.76</td>
<td>1.76</td>
<td>1.158</td>
<td>125</td>
<td>0.249</td>
</tr>
<tr>
<td>See</td>
<td>1.06</td>
<td>1.06</td>
<td>-0.01</td>
<td>125</td>
<td>0.992</td>
</tr>
<tr>
<td>Hear</td>
<td>0.66</td>
<td>0.8</td>
<td>-0.765</td>
<td>125</td>
<td>0.446</td>
</tr>
<tr>
<td>Feel</td>
<td>0.49</td>
<td>0.72</td>
<td>-1.308</td>
<td>125</td>
<td>0.193</td>
</tr>
</tbody>
</table>

To further explore the differences between the words used in filtered versus visible reviews, the review database was sorted from low to high. This approach might allow greater insight into dominant words used in extremely negative or extremely positive reviews, both places where a fake review is more likely to be found. Figure 5 presents all one-star reviews in the restaurant database that were filtered, while Figure 6 presents all one-star restaurant reviews that were not filtered.

Figure 5: Word cloud for all one-star restaurant reviews that were filtered.
While the two word clouds are similar, words about food items, such as "noodles," "oil," "gluten," and "pasta," dominate the filtered reviews. In the visible one-star reviews, service-related words, such as "manager," "served," "order," and "minutes" are more prominent. Again, trigger words, such as "loved" or "hated," are not prominent.

Figure 7 presents a word cloud of all filtered five-star reviews in the religious organizations database, while Figure 8 presents all visible five-star reviews for religious organizations. Again, the word cloud analyses fail to discriminate the differences between filtered and unfiltered reviews. If there are any strong rhetorical differences between the two groups of words, they are not prominent and are buried inside the word clouds.

Summarizing qualitative and quantitative analyses of review content, we have found no explanations to explain review filtering, other than that visible reviews are generally longer than filtered reviews.

Next, turning to reviewer-based variables, Table 7 shows the signals related to the ethos of the people who wrote the reviews in the sample. This analysis shows a clear demarcation between reviewers whose writing is filtered versus those whose isn’t. People who write reviews that Yelp filters have fewer network connections than those whose reviews are visible. In the restaurant review database, the average number of friends for those who write visible reviews is 112, versus seven for those whose reviews are filtered; in the religious organizations database, people who wrote visible reviews have an average of 249 friends while people who wrote filtered reviews have an average of one friend.

<table>
<thead>
<tr>
<th>Table 7: Reviewer signals on Yelp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Restaurant</td>
</tr>
<tr>
<td>Filtered</td>
</tr>
</tbody>
</table>
People whose reviews are visible are more prolific on Yelp; in the restaurant database the average number of reviews written for this group is 210, while the group whose reviews are filtered is only 10. In the religious organizations database, people who authored visible reviews averaged 249 reviews, while people who wrote filtered reviews averaged 11.

Even something as simple as posting an avatar picture makes a difference in how your content is treated by Yelp: for those whose work is visible, 93 percent of restaurant reviewers have posted an avatar photo; for those whose work is filtered, only 35 percent have posted an avatar. In the religious organizations database, 87 percent of the visible reviews were accompanied by an avatar picture, while only 37 percent of the filtered reviews had an avatar. Other signals from the Yelp community also appear to be related to having your work visible: these include having fans and having your work being voted "useful," "funny," or "cool." Visible reviews enjoy many orders of magnitude more social feedback than those that are filtered.

To summarize, people whose reviews are live in the database are prolific reviewers and have built many social signals into their Yelp personas. Generally, they are known and trusted by the community. People whose reviews are filtered have written few reviews and are seen by the Yelp community as isolated and unknown. These results are consistent across two very different categories of reviews on Yelp.

Discussion and conclusion

In this study we have attempted, in a limited manner, to test the question: does Yelp's review filter function as Yelp says it does?

When looking at evidence related to the actual reviews, there appears to be only minor differences between reviews that are filtered and those that are presented in the live database. While we would expect that filtered reviews are likely to be short, in this study they were only a little bit shorter than visible reviews (by an average of 20 words). And, while previous research found that filtered reviews are more extreme, more likely to be one or five stars, in this study, filtered reviews averaged 3.5 stars and non-filtered reviews averaged 3.6 stars in the restaurant database and 4.3 stars and 4.7 stars respectively in the religious organizations database — fairly similar star ratings (Luca and Zervas, 2013). These moderate-to-high averages share similar standard deviations so the measures of central tendency are not masking some unseen extreme behavior in the data.

Further, the word clouds generated by filtered reviews lack any "trigger" words that might betray these reviews as fake. This is seen in all comparisons, whether we look at all reviews, or only one- or five-star reviews. Simply put, the word clouds generated by filtered reviews look very similar to word clouds made by visible reviews.

Quantitative analysis using the LIWC software also failed to discriminate between filtered and visible reviews. If there are differences between filtered and visible reviews, we have yet to find the variables that explain them.

This evidence suggests that the actual content of a review is a minor factor in whether a review is filtered by Yelp.

Moving to social signals that connect the reviewer to the Yelp community, we find strong support for filtering based on the online reputation of the reviewer. People who are strongly linked to the Yelp community through review creation, social interaction and endorsements are unlikely to have their reviews filtered. Individuals who don't write much, aren't endorsed and who don't fill out their user profiles are much more likely to have their reviews filtered.

With these results in mind, it becomes easier to understand why some reviews are filtered: the owner of a new restaurant asks her friends to write some positive reviews to boost business. A disgruntled ex-employee wants to hurt his former employer so he writes a negative review. Or someone writes a negative review of a business based upon the business owner's political orientation. In all of these cases, the writer is not likely to be a trusted member of the Yelp community, so these reviews are largely filtered.

Does Yelp make the wrong call on the filtering of some reviews? Undoubtedly. The size of the review database makes this a certainty. Humans, after all, are still smarter than algorithms. But the wrong calls that are certain to exist must be considered in a proper context. To Yelp, a small fraction of reviews wrongly filtered is an acceptable error. But to a business owner, each review matters very much.

Does Yelp selectively filter reviews based upon the business they're written for? That question is beyond the scope of this study. The data here support the notion that prolific, socially connected reviewers have their reviews published, while people who don't build their network on Yelp and don't write many reviews are much more likely to have them filtered. A look at the review database from the level of an individual business, over time, may yield different results.

Sentiment against Yelp is strong among the business community, and some of that could be mitigated through better customer service. In the complaints to the FTC, there are many frivolous claims. But some appear to be valid. Yelp's inbound communication is poor, and the company appears to be unresponsive about changing out-of-date locations, hours, closed businesses and other time-based issues, such as when a restaurant keeps its name but changes management. In this case, should old reviews be removed? To a business, these are important issues. Ultimately, Yelp must address this trust deficit if the company is to keep its dominant position in review space.

Businesses can lessen their dependence on Yelp reviews by practicing good public relations tactics: by becoming a high-performance company, so the reviews will trend positively, and being responsive to complaints. The Yelp system does allow businesses to reach out to unhappy reviewers, but this should be done with care. Messaging strategies for businesses include building a strong Web site with good search engine optimization. This increases the likelihood that owned media would be the first thing the customer sees. Businesses should also claim their space on other relevant review sites, such as those offered by Google and Bing, and use those to drive traffic to the owned media.

The review space is fast moving and Yelp could change its algorithm in a moment, without notice. For this reason alone, it would be worthwhile to replicate this study. Future studies should work with larger samples, across locations and business categories. Because Yelp is an advertising company, it would be useful to take a direct look at how advertisers and non-advertisers are treated by the company. Indeed, the company should welcome such a study.

About the author

David Kamerer serves as assistant professor in the School of Communication at Loyola University Chicago.
E-mail: david [at] davidkamerer [dot] com

Acknowledgements

The author would like to acknowledge the assistance of Rebecca Browne and Anna Sherman in developing the database used for this study.

Notes

1. Complaints are shown with original spelling, grammar and punctuation.
2. http://www.liwc.net/
References


Editorial history
Received 7 July 2014; revised 16 August 2014; accepted 20 August 2014.

Copyright © 2014, First Monday.
Copyright © 2014, David Kamerer.