Product Reputation Manipulation: The Impact of Shill Reviews on Perceived Quality

Completed Research Paper

Toan Ong  
University of Colorado Denver  
toan.ong@ucdenver.edu

Michael Mannino  
University of Colorado Denver  
michael.mannino@ucdenver.edu

Dawn Gregg  
University of Colorado Denver  
dawn.gregg@ucdenver.edu

ABSTRACT
Online reviews have become a popular method for consumers to express personal evaluations and buyers to use as a source of product quality information in purchasing decisions. However, integrity of review systems is threatened by evidence about the prevalence of shill reviews. To understand the impact of shill reviews, an experiment was conducted to collect shill reviews and measure the impact of shill reviews on perceived product quality. The results showed positive shill reviews significantly increased quality perceptions of consumers for thinly reviewed products. This finding provides strong evidence about the risks of shill reviews and emphasizes the need to develop effective detection and prevention methods.

Keywords  
Reputation manipulation, fake reviews, shill reviews, perceived quality

INTRODUCTION
Buyers use online reviews as a source of knowledge during product purchase decisions. These reviews, reflecting the personal experiences of the reviewers, provide consumers with additional information about a product and allow them to verify that the information advertised by the manufacturer is accurate. Thus, reviews help online buyers make more informed purchase decisions.

Because reviews affect purchase decisions and product sales, sellers have motivation to use fake reviews to mislead buyers with incorrect product information. These fake reviews are “shill reviews”. Lam and Riedl (2004) defines shills as users “whose false opinions are intended to mislead other users”. We extend this definition by specifying that a shill is a person who writes a review for a product without disclosing the relationship between the seller and review writer. Review manipulation was found on online marketplaces like Amazon.com and BarnesandNoble.com (Hu, Bose, Gao and Liu, 2011; Hu, Liu and Sambamurthy, 2011).

The impact of reviews on purchase decisions has been found in multiple studies (Bounie, Bourreau, Gensollen and Waelbroeck, 2005; Chevalier and Mayzlin, 2006). One reason is that consumers use reviews to confirm their perception about product quality (Moe, 2009). Since perceived quality plays an important role in the purchase decision making process (Tsiotsou, 2005; Zeithaml, 1988), reviews can impact customer purchase decisions. Shill reviews can distort perceptions about the quality of target products.

The objective of this study is to understand the impact of positive shill reviews on product quality perceptions. We only consider positive shill reviews in this study because in a large marketplace where each product category has many competitors, positive reviews have more direct impact on the consumers than negative shill reviews. We extended Zeithaml (1988) product quality model to develop hypotheses about the impact of normal and shill reviews on perceived product quality. To test these hypotheses, we designed an experiment using primary data about shill reviews paired with secondary
data about normal reviews. The experiment examines both the impact of the overall product rating on the initial impression of a customer about product quality as well as the impression the customer has about the product's quality after reading normal and shill reviews. An advantage of collecting data via an experiment is the ability to monitor review usage, such as the quantity of reviews read and time spent on reading the reviews, which are difficult to observe. This flexibility supports linkage of review usage behavior and consumer quality perceptions.

Our results demonstrate that shill reviews have a significant effect on product quality perceptions for thinly reviewed products. Thinly reviewed products are susceptible to shill attacks because their rating is easier to improve than heavily reviewed products. The impact of shill reviews on perceived product quality is significant when only the rating summary is used by consumers. When consumers read the content of shill reviews, the impact of shill reviews is even stronger. The results also show that, in general, product reviews significantly impact perceived product quality. Our findings about the effect of shill reviews raises awareness about the risks posed by shill reviews. The findings of this study also provides evidence that word-of-mouth, in online shopping environments, influence perceived product quality.

LITERATURE REVIEW

The effect of product reviews

A review system collects, distributes, and aggregates feedback about products. Literature in economics and marketing has shown that review systems are used widely by consumers to make both online and offline purchases (Bansal and Voyer, 2000; Chatterjee, 2001; Godes and Mayzlin, 2004). Using movie box office data, Liu (2006) and Duan, Gu and Whinston (2008) showed that the quantity of reviews positively impacts movie revenue. On the other hand, a significant effect of average rating of the reviews on product sales was found (Chevalier et al., 2006; Cui, Liu and Guo, 2010; Dellarocas, Zhang and Awad, 2007; Ye, Law and Gu, 2009). Explaining the effect of the rating score on product sales, Forman et al. (2008) stated that consumers use ratings as a measurement for product quality and that a good rating can draw buyers’ attention.

Studies about the effect of review rating score find that the effect of low-end (1,2 star) and high-end (4,5 star) reviews on sales depends on characteristics such as price premium (Chevalier et al., 2006; Clemons, Gao and Hitt, 2006). According to Forman, Ghose and Wiesenfeld (2008), reviews with strong rating (either positive or negative) provide “a great deal of information to inform purchase decision”. However, the rating score is not the only element of the review that has an impact on consumers. Sun (2008) found that even when a product has a low rating score, some of its high rating reviews which create high variance in the review set can increase the sales of that product. The content of the text comment of the reviews influences product sales (Ghose and Ipeirotis, 2004) because consumers want to know what others think about the product (Pang and Lee, 2008).

Review manipulation

There are several approaches to prevent review manipulation. One approach is that websites can encourage more reviews to be submitted making it more difficult to change the average rating (Dellarocas, 2004). To encourage review submission, websites usually give some kinds of rewards to reviewers. Another less common approach is to limit name changes by allowing the user to commit to their identification or charging entry fee (Friedman and Resnick, 2001). An alternative solution, utilized by Amazon.com, is to increase the credibility of the reviews by providing certifications such as publishing the reviewer’s real name or indicating that the review was written by an Amazon verified consumer.

Perceived quality

There are two different kinds of quality: objective quality and perceived quality. Objective quality is defined as “the technical superiority or excellence of the product” (Zeithaml, 1988). Objective quality is the true quality of the product and is often stable (Clark, Doraszelski and Draganska, 2009). Perceived quality is defined as “the consumer’s judgment about a product’s overall excellence or superiority” (Zeithaml, 1988). Perceived product quality is an important factor that impacts consumer

---

1 Multiple websites such as Epinion.com and Ciao.co.uk reward its members for writing reviews. Amazon.com recognizes the effort of posting helpful reviews of the reviewers by creating lists such as Amazon’s Top Customer Reviewers and Hall of Fame Reviewers.
behaviors such as intention to buy or product selection (Jacoby, Chestnut, Hoyer, Sheluga and Donahue, 1978; Sawyer, 1975). Therefore, one way to influence purchasing behavior is to influence perceived product quality. Shill reviews can be used to change the perceived quality judgments of potential customers.

THEORETICAL BACKGROUND

Perceived quality is an important variable in marketing research (Snoj, Korda and Mumel, 2004; Zeithaml, 1988). With the emergence of review systems, a new factor that impacts perceived quality is product reviews. While advertising provides product information from the manufacturer’s perspective, product reviews provide product information from the product user’s perspective. According to a report of Neilson Company in 20092, online opinions are more trusted than most forms of advertising. Li and Hitt (2008) found that prior to buying a product, the product quality expectation of the consumers can be affected by the product reviews. Therefore, we expect that positive shill reviews can impact product perceived quality.

Three metrics, valence, volume and variance, are frequently used to measure the impact of the rating summary on the first impression of product quality. Although the appearance of positive shill reviews in a rating summary is not apparent, all three metrics are affected by the rating of positive shill reviews. Since positive shill reviews usually have high rating (Mukherjee, Liu and Glance, 2012), the appearance of shill reviews in the review set will increase the valence of the product. The volume increases when shill reviews are added. The ratings of shill reviews also increase the variance of product ratings because only negatively rated products need help from shills. So the ratings of shill reviews must be at the opposite direction of other reviews in the review set (Wu, Greene, Smyth and Cunningham, 2010).

Valence is usually represented by the average rating measure (Clemons et al., 2006). Consumers use average rating to compare among products with the average rating serving as a proxy of product quality (Cui et al., 2010). Average product rating has positive impact on perceived product quality (Shen, Li and DeMoss, 2012). Volume is usually measured by the quantity of the reviews. Volume has been found to impact product sales (Awad and Zhang, 2006; Duan et al., 2008; Liu, 2006). Volume of ratings received is one measure to estimate the size of the group of consumers who bought and used a product. A larger group means more people have used the product regardless of the rating they provide. The statistical variance of the ratings represents disagreement among reviewers about a product (Awad et al., 2006). Thus, we hypothesize that:

H1a: The valence of product ratings positively impacts perceived product quality.

H1b: The volume of product ratings positively impacts perceived product quality.

H1c: The influence of variance on perceived product quality is affected by average rating.

Usage of actual review comments can also be related to the valence, volume and variance of reviews. Consumer behavior is affected by perceived risk because any action may have unanticipated consequences (Bauer, 1960). Research shows that average rating is used as a proxy of product quality (Cui et al., 2010; Forman et al., 2008). Thus, better ratings can reduce perceived product risk. Similarly, a large variance shows disagreement among the reviewers which can lead to higher perceived risk (Awad et al., 2006). According to Lutz and Reilly (1974), when perceived product risk is high, consumers tend to collect more information about the product. Review usage can also be related to the number of reviews available to read. Past research has shown that consumers read no more than two pages of reviews (Pavlou and Dimoka, 2006). Since thinly reviewed products have small number of reviews, volume dictates the quantity of reviews available for the consumers to read. Thus, we hypothesize that:

H2: The valence of product ratings negatively impacts review usage.

H3: The volume of product ratings positively impacts review usage.

H4: The variance of product ratings positively impacts review usage.

---

Prior to product purchases, consumers purchasing online usually don’t have physical contact with products. Using the reviews, consumers look for the information from previous product users. By reading reviews, consumers might collect some new information which is not available by using the rating summary only. The new information might change the consumer’s first impression about the quality of the product. As an example, an experiment compared one group of participants that read positive reviews of a film with another group of participants that read negative reviews also of that film (Wyatt and Badger, 1984). After viewing the films, the participants were asked to evaluate the films. The results showed that direction of the reviews significantly impacted direction of the evaluation. We hypothesize that:

H5: The usage of negative normal reviews negatively changes the first impression about product quality.
H6: The usage of positive shill reviews positively changes the first impression about product quality.

MODELS

Three linear regression models were used to test the hypotheses. Model 1 answers the research question about the impact of rating summary on perceived quality. Model 2 measures the influence of rating summary on review usage. Model 3 estimates the impact of review usage on the change of consumer quality perceptions.

Model 1:

$$PQ_i^0 = \alpha_0 + \alpha_1 \text{AveRating}_i + \alpha_2 \text{VolRating}_i + \alpha_3 \text{VarRating}_i \times \text{AveRating}_i + \epsilon_i$$

where:

- $PQ_i^0$ : The first impression of product $i$
- $\text{AveRating} = \frac{\sum_{q=1}^{n_i^N} \text{NormalRating}_{q1} + \sum_{s=1}^{n_i^S} \text{ShillRating}_{qs}}{n_i^N + n_i^S}$
- $\text{VolRating} = n_i^N + n_i^S$
- $\text{VarRating} = \sqrt{\frac{\sum_{q=1}^{n_i^N} (\text{NormalRating}_{q1} - \text{AveRating})^2 + \sum_{s=1}^{n_i^S} (\text{ShillRating}_{qs} - \text{AveRating})^2}{n_i^N + n_i^S}}$
- $n_i^N$ : Quantity of normal reviews in the review set of product $i$
- $n_i^S$ : Quantity of shill reviews in the review set of product $i$

Model 2:

$$TQ_i = \beta_0 + \beta_1 \text{AveRating}_i + \beta_2 \text{VolRating}_i + \beta_3 \text{VarRating}_i + \epsilon_i$$

$$MT_i = \gamma_0 + \gamma_1 \text{AveRating}_i + \gamma_2 \text{VolRating}_i + \gamma_3 \text{VarRating}_i + \epsilon_i$$

where:

- $TQ_i$ : Total number of reviews of product $i$ read by the consumer.
- $MT_i$ : Median time the consumer spends on the shill reviews of product $i$.

Model 3:

$$\Delta PQ = PQ_i^1 - PQ_i^0 = \beta_0 + \beta_1 TQ_i^N + \beta_2 MT_i^N + \beta_3 TQ_i^S + \beta_4 MT_i^S + \epsilon_i$$

where:

- $PQ_i^1$ : Perceived quality of product $i$ after the reviews are read.
- $\Delta PQ$ : Difference in perceived quality before and after the reviews are read.
- $TQ_i^N$ : Total of normal reviews of product $i$ read by the consumer.
- $MT_i^N$ : Median time the consumer spends on the normal reviews of product $i$.
- $TQ_i^S$ : Total of shill reviews of product $i$ read by the consumer.
- $MT_i^S$ : Median time the consumer spends on the shill reviews of product $i$. 

RESEARCH METHOD
Testing the impact of shill reviews on perceived quality requires both shill reviews and normal reviews to be gathered and shown to the subjects. Shill reviews were submitted by shills who have an undisclosed relationship with the seller. A normal review is free of undisclosed relationships between seller and reviewer unlike shill reviews. The collected shill and normal reviews were mixed together to create different review sets. The same product was shown to the subjects along with one of these review sets. Different review sets with different shill and normal review combinations allowed us to measure the impact of the reviews on perceived quality.

Shill review collection
In this study, shill reviews were collected as primary data in which subjects were asked to become shills and intentionally write positive reviews for an MP3 player. Undergraduate students, a convenience population, were chosen for this study because they are active technology and internet users. To simulate real conditions for writing shill reviews, the product information available to the subjects was limited. The subjects were provided with the product specifications and two pictures of the product. To ensure that the reviewers would not seek the product’s information or its reviews online, product identification information such as brand name, product name and model number were changed. The price of the product was also hidden from the reviewer. With the provided information, the subjects were asked to rate the product and write a short review title and a text comment. Although the shill reviewers were asked to submit positive shill reviews, the reviewers were not given specific instructions about review content, structure, and format. The subjects had no specified time limits to write the reviews.

Normal review collection
While the shill reviews were collected as primary data, normal reviews were collected as secondary data on Amazon.com. Although there is no absolute way to verify the lack of undisclosed relationships underlying the reviews collected, risk was reduced by only including reviews that either disclosed the reviewer’s name or were Amazon.com verified purchasers. Shill reviewers will not typically disclose real names in shill reviews because of the risk of losing reputation. It is also unlikely for a shill reviewer to actually buy the product just to have the “Amazon.com verified purchase” badge because it will increase the cost of submitting a shill review.

The experiment
The experiment simulates the situation in which a seller wants to dishonestly promote a thinly reviewed product. In the experiment, an MP3 player was shown to the subjects with basic product technical information along with the overall rating information such as average rating, quantity of reviews and the distribution of the reviews. The review set of the product was a random mix between positive shill reviews and negative normal reviews. With the available information, the subjects were asked to give their opinion about their perception of the product quality. This response could not be changed after submitted. Then, the subjects were shown the rating, title and text comment of all the reviews. The review usage of the subjects was recorded. Finally, the subjects were asked about the product quality perception again. The comparison of the product quality perceptions before and after the reviews were read indicated the impact of the review content on product quality perception.

In this study, unidimensional scale was selected to measure perceived quality because there are too many product quality dimensions mentioned in review content. The review usage was measured in two dimensions: quantity of reviews read and median time spent on each review. The total quantity of reviews read provides the quantity of each type of reviews (e.g. normal and shill) read by the consumers. To isolate the impact of shill reviews, we had to determine that shill reviews were read. Median time spent on a review measures the reading effort of the consumers. To recruit the subjects for this experiment, a probability sample of undergraduates and graduate students was used. 6000 invitation emails were randomly sent to 17,000 students. To comply with the requirement of probability sampling, every student in the list must have the same probability of receiving the invitation email. According to the rules of thumbs suggested by Hair et al. (Hair, Black, Babin, Anderson and Tatham, 2005), the preferred ratio between the sample size and quantity of independent variables is 20:1. The largest model in this study has 4 independent variables. The quantity of students who participated in the experiment was 175.
RESULTS

As indicated in Table 1, there is a significant relationship between average rating and consumer first impression about the quality of the product. The effect size of the model is large. This result supports the statement of Forman et al. (2008) saying that product rating can be used as a proxy of product quality. The results suggest that one additional star in average rating can increase perceived quality rating by 0.72. This result is ideal for shill attack because the direct result of a positive shill review is an improved average rating of the product. Therefore, if consumers use only the average rating to assess the quality of the product, shill attacks can be very effective in making them think that the quality of the product is good. The effects of variance and volume on the first impression were not statistically significant.

<table>
<thead>
<tr>
<th>Model</th>
<th>Coefficients</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>S.E</td>
</tr>
<tr>
<td>Constant</td>
<td>1.935</td>
<td>0.302</td>
</tr>
<tr>
<td>AveRating</td>
<td>0.720</td>
<td>0.074</td>
</tr>
<tr>
<td>VarRating</td>
<td>-0.007</td>
<td>0.059</td>
</tr>
<tr>
<td>VolRating</td>
<td>-0.039</td>
<td>0.031</td>
</tr>
</tbody>
</table>

Table 1 – Factors that impact the first impression

The results from Table 2 show that only volume has a statistically significant impact on the total quantity of reviews read by the consumers. This result makes sense because volume limits the quantity of the reviews that are available for the consumers to read. Furthermore, in this study, the maximum quantity of reviews for each product was 10. Because this quantity is small, the subjects might read entire the review set regardless of the variance and the average rating of the reviews.

<table>
<thead>
<tr>
<th>Model</th>
<th>Coefficients</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>S.E</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.165</td>
<td>0.462</td>
</tr>
<tr>
<td>AveRating</td>
<td>0.147</td>
<td>0.113</td>
</tr>
<tr>
<td>VarRating</td>
<td>0.113</td>
<td>0.090</td>
</tr>
<tr>
<td>VolRating</td>
<td>0.792</td>
<td>0.048</td>
</tr>
</tbody>
</table>

Table 2 – Factors that impact quantity of reviews read

Table 3 indicates that only volume impacts median time spent on each review, but the effect size is very small. When more reviews were available, the subjects spent less time reading the reviews. In combination with previous results, we can conclude that when the quantity of reviews is small, consumers tend to read all the reviews but spend less time on reading each review. 82.43% of the participants read all the reviews that were available.

<table>
<thead>
<tr>
<th>Model</th>
<th>Coefficients</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>S.E</td>
</tr>
<tr>
<td>Constant</td>
<td>25.639</td>
<td>3.347</td>
</tr>
<tr>
<td>AveRating</td>
<td>0.248</td>
<td>0.819</td>
</tr>
<tr>
<td>VarRating</td>
<td>0.037</td>
<td>0.650</td>
</tr>
<tr>
<td>VolRating</td>
<td>-1.054</td>
<td>0.346</td>
</tr>
</tbody>
</table>

Table 3 – Factors that impact median time spent on each review
The subjects were asked to assess the quality of the product at two different periods: before and after they read the content of the reviews. After reading the reviews, 66.9% of the subjects changed their first impression about the quality of the product. The results from Table 4 indicate that shill reviews have a statistically significant impact on the change of quality perception while normal reviews do not. Reading an additional shill review increased the perceived quality rating by 0.095. The results also suggest that one additional second spent on reading a shill review increases the perceived quality rating of the consumers by 0.012.

Further analysis also provides evidence about the effect of shill level on perceived quality. Shill level is measured by the percentage of shill reviews in the review set. There are two different review sets. At rating summary level, the review set contains all shill reviews and normal reviews of the product. After the subjects read the reviews, the review set only contains the shill and normal reviews that are read by the subjects. The results in Table 5 show that the shill level in both datasets has significant impact on perceived quality. On the first impression, if the shill level increases by 1%, the perceived quality rating increases by 0.024. After the subjects read the reviews, this impact is even higher.

**DISCUSSION**

The results suggest that shill reviews are more influential for thinly reviewed products because shill reviews have the greatest potential to improve overall product ratings for products with fewer real reviews. In addition, the effect of shill reviews was even stronger after being read. One explanation for this result is that shill reviews usually are extreme which can be more persuasive to consumers than more neutral (unbiased) reviews (Mukherjee et al., 2012). Another reason for shill reviews to have such an impact is that consumers are unaware of the appearance of shill reviews. These findings should raise the awareness about the danger of review manipulation. Online marketplaces should pay more attention to developing effective shill review detecting methods to protect the consumers and honest sellers.
The empirical results show that consumers use average rating as an indicator of product quality. This result is consistent with the statement of Forman et al. (2008) saying that product rating can be used as the proxy of product quality. Therefore, maintaining a good product rating is a very important task. Regardless of the review quantity, products with a better rating have a better first impression about product quality. However, the review quantity influences the quantity of reviews read by the consumers. When the quantity of the reviews is relatively small (e.g. less than or equal to 10), the consumers tend to read all the reviews.

CONCLUSION
To study the impact of shill reviews on perceived product quality, an experiment was designed using reviews collected from primary and secondary sources. Three linear regression models were used to measure the impact of review rating summary on the first impression and review usage and the impact of review usage on the change in product perceived quality. The results showed that shill reviews have significant impact of perceived product quality. This finding raises awareness about the risks posed by shill reviews. This study is limited only to the usage of positive shill reviews while negative shill reviews can be effective also.

REFERENCES


