

December 2007

An Analysis of the Differential Impact of Reviews and Reviewers at Amazon.com

Samita Dhanasobhon
Carnegie Mellon University

Pei-Yu Chen
Carnegie Mellon University

Michael Smith
Carnegie Mellon University

Pei-yu Chen
Carnegie Mellon University

Follow this and additional works at: <http://aisel.aisnet.org/icis2007>

Recommended Citation

Dhanasobhon, Samita; Chen, Pei-Yu; Smith, Michael; and Chen, Pei-yu, "An Analysis of the Differential Impact of Reviews and Reviewers at Amazon.com" (2007). *ICIS 2007 Proceedings*. 94.
<http://aisel.aisnet.org/icis2007/94>

This material is brought to you by the International Conference on Information Systems (ICIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2007 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

AN ANALYSIS OF THE DIFFERENTIAL IMPACT OF REVIEWS AND REVIEWERS AT AMAZON.COM

Pei-yu Chen

Tepper School of Business
Carnegie Mellon University
Pittsburgh, PA 15213
pychen@andrew.cmu.edu

Samita Dhanasobhon

Heinz School of Public Policy
and Management
Carnegie Mellon University
Pittsburgh, PA 15213
sdhanaso@andrew.cmu.edu

Michael D. Smith

Heinz School of Public Policy
and Management
Carnegie Mellon University,
Pittsburgh, PA 15213
mds@cmu.edu

Abstract

Online product review networks help to transmit information that customers can use to evaluate products in Internet commerce. These networks frequently include an explicit social component allowing consumers to view both how community members have rated individual product reviews and the social status of individual reviewers. We analyze how these social factors impact consumer responses to consumer review information. We use a new dataset collected from Amazon.com's customer reviews of books. This dataset allows us to control for the degree to which other community members found the review helpful, and the reputation of the reviewer in the community. We find that more helpful reviews and highlighted reviews have a stronger impact on sales than other reviews do. We also find that reviewer information has a stronger impact on less popular books than on more popular books. These results suggest that the dynamics of reputation communities make it harder for self-interested parties to manipulate reviews versus an environment where all reviews are treated equally.

Keywords: Electronic Commerce, Recommendation System, Digital Word-of-Mouth

Introduction

Digital networks for product information have redefined traditional “word-of-mouth” social networks by allowing consumers to easily share their opinions and experiences with other members of large-scale online communities (Dellarocas 2003). Many online retailers, such as Amazon.com and BarnesandNoble.com, are augmenting their product markets by building online communities to provide product reviews to other consumers. Likewise, many auction sites, such as Ebay.com, allow consumers to rate product sellers. Such information sharing has the potential to reduce the uncertainty consumers face regarding the quality of a product or a seller.

Several papers in the literature have shown that large-scale information sharing in digital networks may help communicate product/seller quality and build trust between buyers and sellers in online markets (Ba and Pavlou 2002; Resnick and Zeckhauser 2002; Dellarocas 2003; Chen and Wu 2005; and Chevalier and Mayzlin 2006). For example, Resnick (2002) shows that seller reviews in eBay influence the probability of a sale, while Chevalier and Mayzlin (2006) find that product reviews at Amazon.com impact book sales. Dellarocas (2005) shows that strategic manipulation of consumer reviews can either increase or decrease the information value of a review to consumers.

This raises the question of why consumers would trust the information provided by strangers they may have never met and how trust is formed among consumers themselves. Credibility is a critical issue in effective information sharing, which involves information reliability and consumer trust. There is an extensive literature in the field of social psychology that shows the importance of credibility in influencing the impact of a persuasive message, where credibility can be based either on the reputation of the author or the content of the message (Cialdini 2000). Indeed, in offline social networks, consumers usually attach different weights to different information sources according to their social-ties and knowledge about the source and the information.

Thus, to realize the full potential and benefits of information sharing in online community, it is essential to have an effective mechanism to help consumers gauge information reliability and to enhance consumer trust. To this end, many retailers have invested in rating systems that allow consumers to provide and read reviews not only on the product *per se*, but also on the credibility of the review message and the reviewer. For example, Amazon.com not only lets its customer post reviews of products, it also allows customers to vote on whether posted reviews were helpful to them in making a purchase decision. The proportion of helpful votes a review receives can serve as an indicator for the quality of the reviews to other consumers (content quality). Furthermore, Amazon.com identifies individual reviewers based on a ranking system where reviewers who post more reviews and have a higher number of helpful votes are singled out to other community members (reviewer quality).

Most of the existing empirical literature on online word-of-mouth focuses on aggregate numerical review scores. However, while providing an important contribution, this only tells part of the story. As noted by Resnick et al. (2000): “these simple numerical ratings fail to convey important subtleties of online interactions. For example, ... what were the reputations of the people providing the feedback?” It remains an open empirical question as to what extent community evaluations of individual reviews and individual reviewers influence consumer purchase decisions online. Our research aims to bridge this gap by considering the role of information credibility in consumer decision-making along two dimensions: the quality of the content and the quality of the source of information.

To do this, we use consumer reviews for books sold at Amazon.com to analyze how the content of reviews and the reputation of the reviewer impact consumers’ responses. We selected Amazon.com, because it is one of the largest online retailers and has one of the most active reviewing communities online (Chevalier and Mayzlin 2006). Within Amazon’s online review communities, we consider three primary measures of information credibility. First, we consider the quality of the content, which can be indexed by how helpful the community found the review (Amazon lists “x of y people found this review helpful,” see Figure 1). Second, we consider the quality of the information source, i.e., the reputation of the reviewer. Amazon identifies its “top” reviewers (see Figure 1), and there is anecdotal evidence that these popular reviewers can have a large impact on book sales (e.g. Paumgarten 2003). Third, we analyze the impact of spotlight reviews when they are displayed on the product page (see Figure 2). These “spotlight” reviews are set apart from the other reviews and are shown before other reviews, so they may have a relatively stronger effect on book sales than other reviews do.

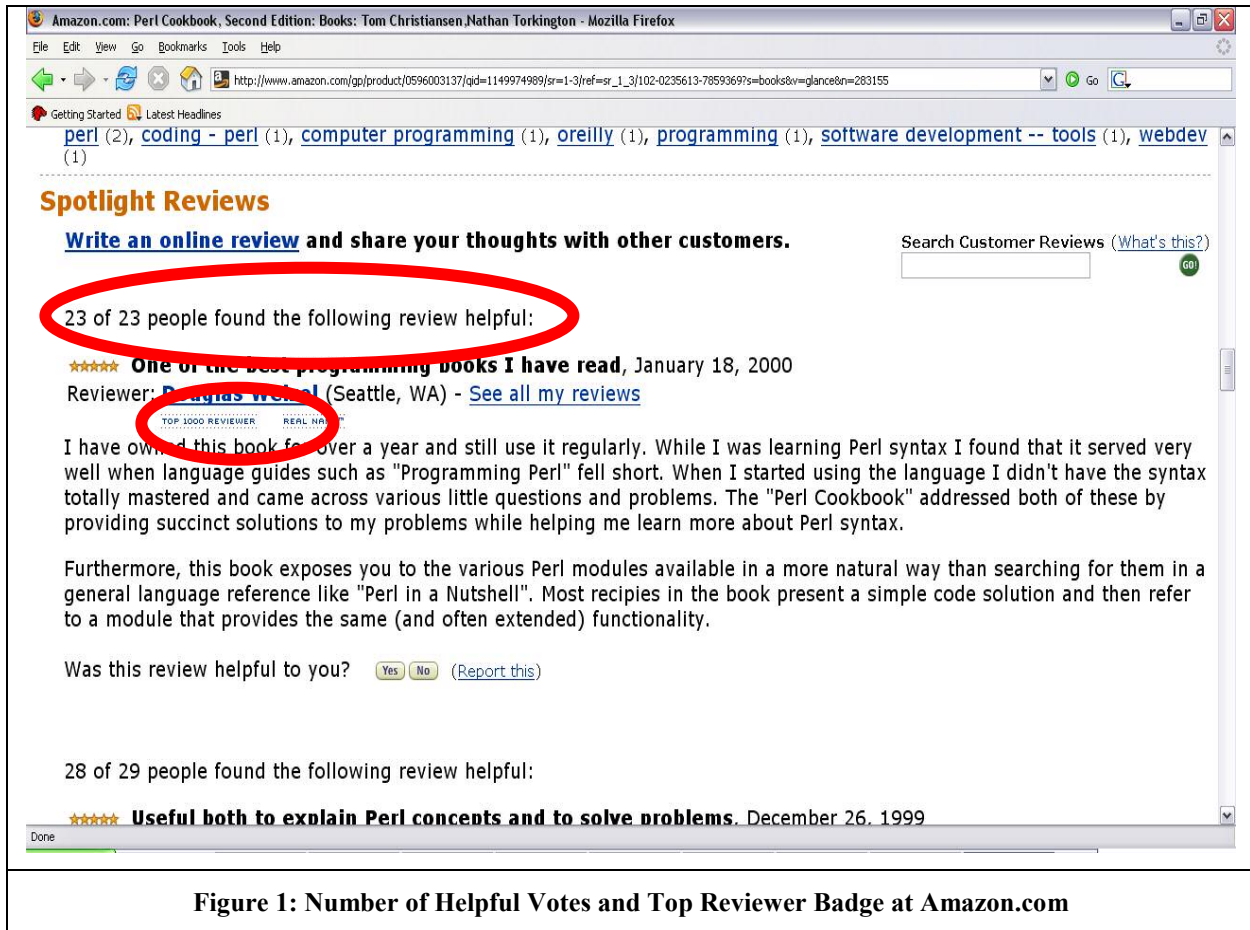


Figure 1: Number of Helpful Votes and Top Reviewer Badge at Amazon.com

To analyze these questions, we collect data daily on product sales levels and customer reviews from Amazon.com's web marketplace. Our data include 50,626 observations of 535 newly released book titles, collected over 195-day period from November 11, 2005 to May 25, 2006. In addition to confirming Chevalier and Mayzlin's (2006) prior finding that higher average star ratings of books are associated with higher sales, we find that reviews with a high proportion of helpful votes (i.e., quality reviews) and spotlight reviews are both associated with increasing sales even after controlling for average star ratings. Further we find that this review information has a stronger impact on less popular books than on popular books. Finally, we find no evidence that the social status of reviewers has an impact on consumer purchase behavior. Together these results suggest that consumers associate different weights to the different messages they receive in making purchase decisions, and reviews are more important for consumers when less outside information is available on the product (as in the case of less popular books).

This research makes two unique contributions. First, this study uses real data in an online setting to unpack the mechanisms that drive people to trust and respond to online product reviews. To the best of our knowledge, this has not been studied empirically in the literature. Second, this research allows us to understand what form of micro-level dynamics in community interactions may be valuable in signaling quality, over and above aggregate-level summary quality scores. This has important implications. When only the aggregate measures are available (i.e. where an uninformative review from a new community member carries as much weight as an informative review from an established and respected community member) it may be easier for self-interested parties to manipulate review results. However, if the micro-level dynamics of reputation communities are important factors in determining product sales, it would be harder for self-interested parties to manipulate reviews, making the reputation system more reliable. Thus, understanding the micro-level dynamics of virtual communities has important implications for designing a more reliable reputation system that is less subject to manipulation.

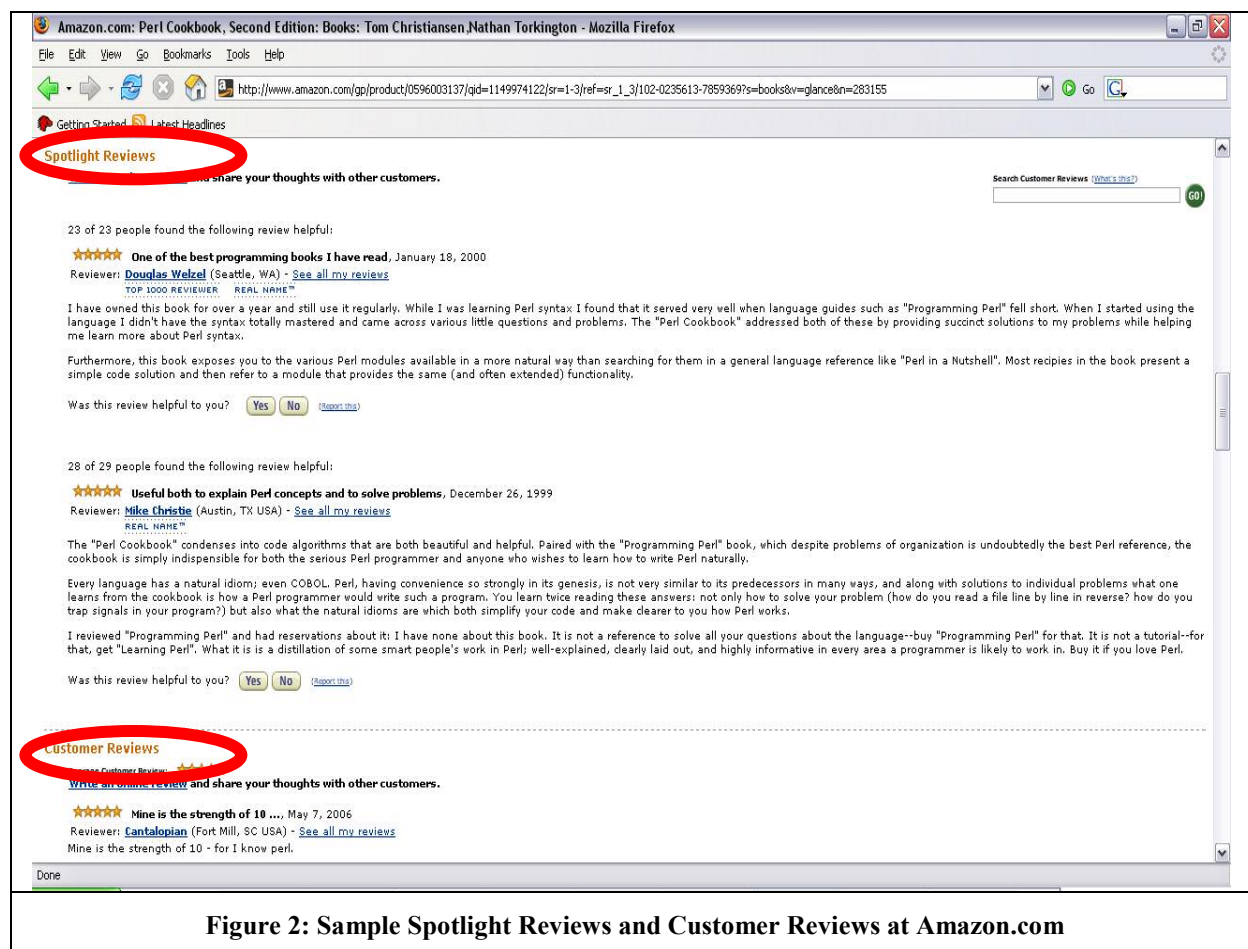


Figure 2: Sample Spotlight Reviews and Customer Reviews at Amazon.com

The remainder of this paper proceeds as follows. In the next session, we review the prior literature as it relates to our research setting. Then, we use this literature to develop our research framework and hypotheses. We present our data and model specifications in data and methodology section respectively. We present our empirical results in the results section and then conclude and discuss the implications of our findings.

Literature Review and Research Framework

In digitally mediated markets, absent reputation systems, consumers can face higher uncertainty about product quality, since there are fewer quality and trust cues available than what is possible in brick-and-mortar markets. To compensate for the lack of quality and trust cues in online markets, many retailers provide rating systems for consumers to rate products and/or allow consumers to write reviews about the quality of products. Such information sharing among consumers provides the potential for reviewers to reduce consumer uncertainty about product quality (Dellarocas, 2003). The search literature has shown that better information on product quality has an impact on consumers' purchase decisions (Stiglitz, 1989).

The prior literature has also shown that when product quality cannot easily be verified, the reputation of the supplier of the product may be used by consumers as an indication for the quality of the product (Resnick et al. 2000). As a result, the perceived quality of the product supplier/seller will have an impact on consumers' purchase decisions and the resulting product sales. This is especially true for auction sites, where information sharing on particular products is less relevant because each "product" is a combination of product and seller, and thus is essentially different.

However, Jin and Kato (2004) noted that consumers must be aware of the posted messages about product supplier/seller quality because the cost of switching identity is very low and the supplier/seller can cheat the systems by manipulating their own profiles. Dellarocas (2005) stated that traditional feedback mechanisms are either automatically recorded or explicitly self-reported. In order to prevent potential fraud from explicit feedback, implicit

feedback (such as Google's PageRank) and distributed reputation mechanisms were created. By using distributed reputation mechanisms, consumers will use both their neighbors' recommendation weighted by the relative amounts of trust, and their own subsequent experience with the sellers to decide whether to buy products from those sellers.¹

Furthermore, the social psychology literature has shown that consumers' purchase decisions are influenced by their attitude toward a product, which may be influenced by the external messages they receive, and that the magnitude of this influence depends on six basic principles of influence on persuasion and attitude change: reciprocity, commitment and consistency, social validation, liking, authority, and scarcity. Overall, this literature suggests that credibility influences the impact of a persuasive message, where credibility can be based either on the reputation of the information source or the content of the message in itself. Note that there is a subtle difference between the reputation of a review supplier and the reputation of a product supplier, because the product supplier is direct beneficiary of high product sales, while the review supplier is not necessarily a direct beneficiary of high sales.

Most of the empirical literature in Information Systems, economics and marketing focuses on the relationships between reviews on products and sales, reviews on retailers and sales, and reviews on individual sellers and sales. In this literature, Resnick and others have shown that reviews of product suppliers represent a good proxy for the reputation of product supplier, and have an impact on product sales characteristics such as the price premium and the probability of a sale. Findings on the relationship of product reviews and sales are mixed, with Chevalier and Mayzlin (2006) showing that an improvement in the review score of a book leads to an increase in relative sales at Amazon.com, while Chen and Wu (2005) and Duan et al. (2005) show that high product ratings do not necessarily lead to increased sales.

With regard to the empirical research in the social psychology literature, Guadagno and Cialdini (2003) note that only two of six principles — authority and consistency — have been examined empirically in an online context. They further observe that status is a meaningful social category in online markets and can translate into higher compliance, particularly when the influence agent is a high-status in group member. Guadagno and Cialdini (2003) also note that online influence attempts may or may not function similarly to attempts in other contexts due to many factors, such as the nature of influence attempt (interactive or static), the amount of prior exposure between message sender and receiver, and the status of message sender in the receiver's group.

Thus, we believe research is needed to address the impact of social cues in review communities on product sales. This paper aims to bridge this gap in the literature by developing a theoretical model and empirical results that not only study the impact of product/seller reviews on sales but also unpack the micro-level impact of consumer interactions on sales and how credibility might be built in the online context. Specifically, since Amazon.com shows average book review ratings to the consumers, we use average book review ratings to control for product quality. We use reviews that receive high helpful votes by other consumers and spotlight reviews to indicate the quality of reviews (i.e., content quality) that may be used by consumers to form their purchase decision. Finally, we measure the reputation of reviewer (i.e., information source) by their standing in the Amazon.com community as ranked by Amazon.com. We identify that top reviewers have their rank below 1000 because Amazon.com only shows special badges for top 1000 reviewers (Figure 1).

Formally, consider the following model: A consumer's decision to purchase a book is influenced by the quality of the book, since the book is an experience good. Therefore, it is difficult to determine the real quality of the book, and instead the consumer must base her decision on the measured quality, which is a function of her prior belief about the book and other cues she has pertaining the quality of the book. That is,

$$\bar{q} = \alpha \bar{q}_p + (1 - \alpha) \bar{q}_M \quad (1)$$

where \bar{q} is the derived quality measure for the book, \bar{q}_p is consumers' prior about the quality of the book, α is the weight a consumer puts on her prior while $1 - \alpha$ is the weight on other available information, and \bar{q}_M is the measured quality based upon all other information.

The purpose of this paper is to examine how consumers form \bar{q}_M , given \bar{q}_p and α . So, we now shift our focus to the determination of \bar{q}_M . Giving a set of N messages (or ratings/reviews) available to the consumers:

¹ Dellarocas (2005) cited Despotovic and Aberer (2004).

$r_1, r_2, r_3, \dots, r_N$, based upon previous theory literature on economics and social psychology, we can construct the influence of these messages pertaining the quality of a book to a consumer in the following function:

$$\bar{q}_M = \sum_{i=1}^N r_i w_i \delta_i \quad (2)$$

where w_i is a measure of the reputation of the reviewer who writes review i , while δ_i is the measure of the quality (or trustworthiness) of review i . Note that while it is possible that a reviewer with higher reputation (i.e., high w_i) may write reviews that are of higher quality (i.e., δ_i), we do not enforce this constraint but leave it to the data to decide.

When consumers consider only reviewer reputation in making decision (that is, they trust the messages a reputed reviewer says regardless of the content) we have

$$\bar{q}_R = \bar{q}_A + \sum_{i=1}^N r_i \left(w_i - \frac{1}{N} \right) \quad (3)$$

where \bar{q}_R is the derived quality measure weighted by reviewer reputation.

When consumers consider content quality (i.e., the trustworthiness of a message) regardless of its source, then the quality measure is weighted by content quality.

$$\bar{q}_C = \bar{q}_A + \sum_{i=1}^N r_i \left(\delta_i - \frac{1}{N} \right) \quad (4)$$

In the case where there is no information available on content quality and reviewer reputation, or information on content quality and reviewer reputation is not used (this is the case for most extant empirical studies on online markets), the most reliable and unbiased index a consumer can use to signal for the quality of the book is the average of all ratings provided by all reviewers, i.e.,

$$\bar{q}_A = \bar{r} = \frac{\sum_{i=1}^N r_i}{N} \quad (5)$$

where \bar{q}_A stands for the aggregate measure of quality taking each review equally.

The contribution of this paper is to bring in the importance of content quality and reviewer reputation and construct a model that allows us to test the role of content quality and reviewer reputation in influencing consumer decisions after controlling for the average ratings.

Given equations (1) and (5), we have:

$$\bar{q} = \bar{q}_A + \sum_{i=1}^N r_i \left(w_i \delta_i - \frac{1}{N} \right) \quad (6)$$

When any $w_i \delta_i, \forall_i \in \{1, 2, \dots, N\}$ deviates from $1/N$, it indicates that consumers are more sophisticated and take into account the content quality and reviewer reputation. We can further test the individual impact of content quality and reviewer reputation by examining equations (3) and (4) to consumer decision making and by assuming the other has no impact (i.e., taking the value 1). For example, by examine the relationship between

$\bar{q}_R = \bar{q}_A + \sum_{i=1}^N r_i \left(w_i - \frac{1}{N} \right)$ and consumer decision, one can show that if any $w_i, \forall_i \in \{1, 2, \dots, N\}$ deviates

from $1/N$, it follows that reviewer reputation does matter in consumer decision making. Moreover, by examining

the relationship between $\bar{q}_C = \bar{q}_A + \sum_{i=1}^N r_i \left(\delta_i - \frac{1}{N} \right)$ and consumer decision, one can show that if any $\delta_i, \forall i \in \{1, 2, \dots, N\}$ deviates from $1/N$, it means that content quality matters in consumer decision making. These deviations can be measured by the coefficient measures in the empirical model by including the proper variables of interest.

In the next section, we apply this framework to develop the theoretical hypotheses we test in this study.

Hypotheses

As noted above, the goal of this paper is to examine how the quality of online product reviews and the reputation of online reviews impact how customers make purchase decisions in the presence of a large-scale word-of-mouth review system. To this end, we develop the following hypotheses.

H1 (Average Rating Hypothesis): Higher product ratings are positively associated with higher sales.

Aggregate measure of product ratings corresponds to \bar{q}_A in previous section. Higher ratings of a book convey two signals: they may indicate high quality of the book, and they also imply that the general public like the book. According to the social psychology literature (Cialdini 2000), social validation has an influence on consumers' attitude. Accordingly, consumers may be more willing to purchase a book that has acquired a social validation than one without, *ceteris paribus*. As noted above, this result has been studied in the prior literature with Chevalier and Mayzlin (2006) finding that higher product ratings lead to higher sales in the context of Amazon.com. We develop this hypothesis in order to confirm the results of Chevalier and Mayzlin (2006).

H2 (Content Quality Hypothesis): Reviews with a high proportion of helpful votes have a relatively higher impact on sales than reviews with a low proportion of helpful votes do.

In a review system, a moral hazard problem can occur where reviewers post unreliable reviews. Likewise, an adverse selection problem can occur because consumers do not know if posted reviews are reliable until they decide whether to buy the product or not. We suggest that the provision of the quality of the review content can be used as a sanctioning device to alleviate moral hazard problem and a signaling device to alleviate adverse selection problem. Hypothesis H2 aims to test the effect of the provision of content quality, i.e., whether δ_i deviates significantly from $1/N$, which corresponds to the coefficient measure of content quality after controlling for average ratings (\bar{q}_A). The perceived reputation or quality is traditionally derived from explicit feedback from transaction participants (Dellarocas 2005). At Amazon.com, after consumers read a review, they can express the helpfulness of the review by voting "Helpful" or "Not helpful." Reviews that have high proportion of helpful votes vouch for the quality of the review by indicating that the community validates those reviews. If consumers observe that reviews receive a very low proportion of helpful votes, they will learn that those reviews are unreliable and they will update their beliefs and weights of each review before they make a decision about purchasing.

H2a: High star ratings of reviews with high proportion of helpful votes results in an increase in sales.

Once the positive reviews receive high proportion of helpful votes, those reviews can motivate readers to purchase the item because many consumers support what reviewers like about that item. On the contrary, when the negative reviews receive high proportion of helpful votes, a lot of consumers believe that the product is not good and other consumers should not buy that product. Hence, the negative reviews that receive a high proportion of helpful votes can result in a decrease in sales.

H2b: Low star ratings of reviews with high proportion of helpful votes results in a decrease in sales.

H3 (Reviewer Reputation Hypothesis): Reviews by more prominent members of the community have higher impact on sales than reviews by other consumers.

This hypothesis aims to test whether w_i deviates significantly from $1/N$, which corresponds to the coefficient measure of reviewer quality after controlling for average ratings (\bar{q}_A). We use the rank of the reviewers as a proxy for the reputation of the reviewers. The ranking of reviewers is determined by Amazon using a combination of the quantity and the quality of reviews submitted. Amazon.com takes into account the popularity of the item being reviewed when tabulating the "helpful" votes. A reviewer who reviews only best-selling items is going to receive more votes than a reviewer who takes only more obscure items into consideration, so they treat these types of reviewers equitably in determining reviewer rank. The rank of "top reviewers" is identified next to the reviewer's name in the review display, and thus is readily available to customers viewing the reviews. We identify top 1,000 reviewers as prominent members because Amazon shows a rank badge next to any top 1,000 reviewers (Figure 1). Cialdini (2000) identified authority or reputation as one of the basic influential principles. Chaiken and Eagly (1983) suggested that the cues of communicator are less important than the characteristics of the message content. While Dubrovsky, et al. (1991) showed that status and expertise are less significant in the computer-mediated decision groups than in the face-to-face interaction, Guegen and Jacob (2002) showed that status and expertise create higher compliance, especially when the message came from a high-status member. Guadagno and Cialdini (2003) summarized these results as "Authority is successful in increasing compliance in online groups when it is used as a decision heuristic, but is far less influential when present in an interactive discussion." Since online product reviews are non-interactive, we hypothesize that more prominent reviewers (reviewers who have higher status member) have higher influence over consumer decisions.

H4 (Spotlight Review Hypothesis): Spotlight reviews have a larger positive marginal impact on sales than other reviews do.

Amazon.com shows the two "best" reviews before others, and highlights them as "Spotlight" reviews. As more helpful votes are cast on individual reviews, the "Spotlight" reviews are updated to reflect the most recent voting. Note that since spotlight reviews are easier to access, they may carry higher weight in consumers' decision making, i.e., they are likely to change consumers' weight distribution on δ_i . Prior research in online markets has shown that consumers perceive relatively high costs associated with processing information online (Brynjolfsson et al. 2005) and that the order information displayed to the consumer has a disproportionately strong impact on their behavior (Smith and Brynjolfsson 2001). Both of these results are reasonable in our setting given that consumers who have limited time may spend relatively more time reading reviews displayed first in the list of reviews. Because of this, we hypothesize that, *ceteris paribus*, spotlight reviews will have a larger impact on sales than other reviews do.

H5: Reviews with high proportion of helpful votes have a larger impact on less popular books than on more popular ones.

This hypothesis aims to test about the impact of consumers' prior beliefs of product quality on her decision making. When consumer has a more confident prior, α is likely to be higher, and \bar{q}_M will therefore be less important in the consumers' decision making. Reviews for more popular books (for example New York Times Bestsellers), are readily available to consumers through other channels (e.g., book clubs, newspapers), and thus consumers may come to Amazon with a strong prior belief about the quality of these books. For less popular books, particularly newly released titles, consumers will have fewer quality cues to rely on, and therefore may place a higher weight on the reviews available at the Internet retailer. Because of this, we hypothesize that reviews will have a larger impact on less popular books than on more popular books.

Data

Our data are collected from publicly available information on Amazon.com's marketplace. Data were collected using Perl scripts to parse data from the relevant HTML pages and, where possible, from the XML data feed

Amazon.com provides to its developers. Our data consist of 535 (20 titles per week x 28 weeks) new books released over a 195 day period from November 11, 2005 to May 25, 2006.² We focus on newly released titles, because consumer opinions are less well formed for these products, making the product reviews more important for consumer purchase decisions. There are also significant changes in the number of reviews for these titles (initially zero, increasing over time), providing an additional source of variation in our data.

To create our sample of books, we first collect the list of all upcoming book releases as listed by Buy.com. We randomly select 20 unique titles from the list of titles in each week. For each title, we begin collecting data on the first day the book is released.³ For this sample, we extracted generic information of each book, such as its International Standard Book Number (ISBN), title, author, release date, and category from Amazon.com. The categories of the books are summarized in Table 1. This data should be constant across sites for a particular book, and we collected it from Amazon.com out of convenience. In addition, for each book in our sample we collected daily information from Amazon.com on the price, sales rank, and the time until the book would ship.

Table 1: Category of Books		
Category	Freq.	Percent
Adult Fiction	109	20.57
Adult Non-Fiction	41	7.74
Do It Yourself	17	3.21
Entertainment	14	2.64
Juvenile	81	15.28
Language & Arts	51	9.62
Professional	74	13.96
Self-Improvement	63	11.89
Social Science	63	11.89
Travel	17	3.21
Total	530	100

Following the literature (Chevalier and Goolsbee 2003; Brynjolfsson et al. 2003), we use the sales rank listed at Amazon.com as a proxy for product sales.⁴ Sales rank is the ordinal ranking of sales of a product within its product category, with the #1 ranked product having the highest sales. Thus, sales and rank move in opposite directions.

Prior work has shown that the relationship between sales rank and sales follows a Pareto distribution:

$$Quantity = \beta_1 Rank^{\beta_2} \quad (7)$$

This relationship can be parameterized either by direct observation of sales levels and resulting sales ranks for a number of titles, data that typically is available from Amazon's suppliers (Brynjolfsson et al. 2003), or by means of an experiment (Chevalier and Goolsbee 2003). Lacking direct supplier data, we used the experiment proposed by Chevalier and Goolsbee to parameterize this relationship, yielding a slope parameter of $\beta_2 = -0.954$.⁵ This estimate is in the range of coefficient values reported by other studies in the literature (albeit at the high end of this range).⁶

² Experimentation with different temporal sub-samples shows that our results are not sensitive to seasonality.

³ Amazon.com does not allow consumers to post reviews before the book is released, so beginning to collect data prior to release would not provide any additional information for the purposes of our study.

⁴ This technique has also been applied in a variety of other studies, including Chevalier and Mayzlin (2004); Ghose, Smith, and Telang (2006); and Ghose and Sundararajan (2005).

⁵ We conducted our experiment on February 14, 2006, by ordering 7 copies of two book titles from different buyer accounts. We picked two book titles that had steady movement in ranks for six months, and tracked the movement

Finally, we collected daily information regarding the reviews posted at Amazon for each book in our sample. For each book review, we collected the review's posting date, the full text of the review, the 1 to 5 star rating given in each review, the identity of the reviewers, whether the reviewer was identified as a "top" reviewer (see Figure 1),⁷ the number and proportion of helpful votes (see Figure 1), and whether the review was highlighted as a "spotlight" review (see Figure 2).⁸ Table 2 provides summary statistics for our data.

Table 2: Summary Statistics					
Variable	Obs	Mean	Std. Dev.	Min	Max
No. of days since release	13,783	62.54	44.66	0	188
Sales rank	13,783	479,534.1	730,078.8	4	3,934,116
Amazon Price	13,783	18.52	20.26	3.99	199
Number of reviews	13,783	9.7	16.47	0	170
Average star rating of all reviews	13,783	4.43	0.702	1	5
Average star rating of reviews that have more than 80% helpful vote	10,798	4.6	0.65	1	5
Average star rating of reviews by top 1000 reviewers	5,830	4.59	0.55	2	5
Average star rating of spotlight reviews	4,750	4.44	0.61	1	5

We note that the number of observations of sales rank and price are less than the number of observations of days since release because the rank and the price of some books were temporarily removed when the books were out of stock. The number of observation of Amazon star rating is less than the number of observation of the 'number of reviews' because several observations have no reviews.

Methodology

We adopt the same first-difference strategy used in Chevalier and Mayzlin (2006), while incorporating measures for the quality of the content of the reviews and the standing of the reviewers in the community. Chevalier and Mayzlin define the book's sales rank as a function of a book's fixed effect (v_i) and other factors that may impact the sales of a book, and use a constant elasticity demand specification.

Specifically, to study the impact of reviews and the quality of reviews on sales, we consider the following model:

$$\ln(rank_i^t) = v_i + \alpha \ln(P_i^t) + \Pi S^t + \Gamma R^t + \varepsilon_i^t \quad (8)$$

where $rank_i^t$ is the Amazon sales rank of book i at time t ; P_i^t is the Amazon price of book i at time t , and α is the own-site price effect; S is a vector capturing the shipping times promised for book i , and Π captures the effect of the shipping time; v_i is a book fixed effect, summarizing the impact of other (unobserved) variables, such as the

of rank for 24 hours after we bought those items. The rank of one book title jumped from 662,973 to 5,521 and the rank of another title jumped from 868,303 to 5,529.

⁶ These coefficients include -0.834 (Weingarten 2001), -0.855 (Chevalier and Goolsbee 2003), -0.871 (Brynjolfsson, Hu, and Smith 2003), -0.877 (Ghose, Smith, and Telang 2006), and -0.952 (Poynter 2000). Using a lower coefficient value (for example -0.834) would only affect our price elasticity result, not our main findings.

⁷ Top reviewers are selected by Amazon based on the number of reviews they post across all products on Amazon's site and the number of helpful votes they receive from other community members for their reviews.

⁸ As noted above, Amazon.com highlights two reviews for the "spotlight" position at the top of the review listings based on the number of helpful votes assigned to that review.

inherent popularity of the book subject, the author, and unobserved marketing variables that contribute to book sales; R is a vector summarizing review activities and Γ measures its impact; ε_i^t captures random effects summarizing all other unknown variables; and t is the number of days since the book was released.

It is important to note that books are typically printed in large quantities prior to being introduced to the market and are sold to downstream retailers at a marginal price that is independent of the quantity sold (Brynjolfsson, Hu, and Smith 2004). Because of this, we follow the prior literature in this area (Chevalier and Goolsbee (2003), Ghose et al. (2006), and Chevalier and Mazylin (2006)) and do not consider price and quantity as jointly determined in our models. This also raises the question of identification. We also note that our analysis of Amazon's pricing patterns suggests that the identification of our models comes from pricing experiments that Amazon may be conducting over time to determine the price sensitivity of their customers.

Taking the difference of equation (8) at time t and time 0 allows us to eliminate the unobserved book fixed effect (v_i) and provides us with the specific model we estimate:

$$\Delta r_i^t \equiv \ln(rank_i^t) - \ln(rank_i^0) = \alpha(\ln(P_i^t) - \ln(P_i^0)) + \Pi(S^t - S^0) + \Gamma(R^t - R^0) + \varepsilon_i^t \quad (9)$$

Within this model, we use different measures of R to fit the model. In our base model we use the number of reviews and the average star rating across all reviews. We extend this to measure information quality by adding variables for the average star rating of reviews with a high proportion of helpful votes, the average star rating of top reviewers, and the average star rating of spotlight reviews. We define the average star rating of the high proportion of helpful votes as the average star rating of the reviews that have more than 80% of helpful votes. The average star rating of top reviewers is the average star rating of reviewers who are ranked by Amazon in the top 1,000 reviewers on the site. The average star rating of spotlight reviews is the average star rating of reviews in the spotlight review section.

Results

We now fit these empirical models to our data. Our correlation matrix is shown in Table 3 and our results are shown in Tables 4 and 5.⁹ In these results, our control variables are consistent across specifications and have the expected signs. The coefficient on price is positive and significant across all specifications meaning that, as expected, when price rises, sales rank rises and sales fall. Multiplying these coefficients by the sales-rank coefficient β_2 , estimated above, yields an own price elasticity in the range of -0.5249 to -0.8132, which is in the middle of the range of own price elasticity for Amazon found in prior studies (Chevalier and Goolsbee 2003; Ghose, Smith, and Telang 2006).¹⁰ The coefficient on the number of days since release is positive, suggesting that these books follow a normal sales lifecycle of declining sales over time. The coefficient on "ships within 24 hours" suggests that faster shipping is associated with higher sales compared to longer shipping times. Finally, the coefficient on the log of the number of reviews is positive which, because this is a first-difference model, suggests that the marginal impact of an additional review declines with the number of reviews, a result that is also consistent with expectations.

With respect to our review variables of interest, we find that higher overall star ratings have a positive impact of sales (negative impact on sales rank). This is consistent with Chevalier and Mayzlin's (2006) prior findings in the context of Amazon's marketplace and is consistent with Hypothesis 1.

In Model 2 of Table 4, we add a variable for the average star rating among reviews with more than 80% helpful votes. The coefficient on this variable is negative and significant in all model specifications. This finding suggests that reviews that are identified as helpful make additional contributions to sales beyond average star ratings (which take into account every review posted on the site and weighs each review equally). Specifically, Model 2 of Table 2 shows that, while average star ratings are associated with higher sales (-0.375), high ratings of quality reviews are associated with an additional marginal increase in sales (-0.5). Overall, this finding shows that, consistent with

⁹ We note that R-Squares in our models are quite low, but Chevalier and Mayzlin (2006) also report R-Square between 0.037 – 0.236 in their difference-in-difference models.

¹⁰ Chevalier and Goolsbee (2003) report own price elasticity of -0.45 and Ghose, Smith, and Telang (2006) report own price elasticity of -1.16 and -1.40.

hypothesis 2, quality reviews (i.e., reviews with high helpful votes) are associated with higher sales than other reviews are, suggesting that consumers attach more weight to these quality reviews in making purchase decisions.

In Model 3 of Table 4, we include a variable for the average star rating among reviews from the top 1,000 reviewers at Amazon.com. Reviews from these individuals are specifically flagged on in the review listings (see Figure 1), and thus might have a larger impact on consumer behavior. However, in our results the coefficient on this variable is small and statistically insignificant. Thus, we fail to accept Hypothesis 3 that reviews from more prominent members of the community will be more influential than other reviews. This result may imply that customers do not trust top rank reviewers as much as we expected, or it may imply that to become a top reviewer you have to review so many products that you can't show the specific product expertise that is expected by the community. We discuss this finding in more detail in the discussion section.

Table 3: Correlation Matrix								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sales rank (1)	1							
Amazon Price (2)	0.113	1						
No. of days since release (3)	0.0628	-0.0377	1					
Shipping (4)	-0.0994	-0.139	0.0807	1				
No. of reviews (5)	0.1504	-0.0355	0.173	0.1643	1			
Average star rating of all reviews (6)	-0.1592	0.0284	-0.1339	-0.0982	-0.4493	1		
Average star rating of reviews with more than 80% helpful vote (7)	-0.1502	-0.0612	-0.0002	-0.0709	-0.027	0.0892	1	
Average star rating of reviews by top 1000 reviewers (8)	-0.0726	0.0136	-0.0388	-0.0785	-0.1236	0.2265	0.2956	1
Average star rating of spotlight reviews (9)	-0.2227	-0.0942	0.0413	-0.1701	-0.2888	0.3484	0.6973	0.4299

Since the reviews from top reviewers are not significant in Model 3, we remove that variable from the model and add spotlight review variable into Model 4. The result in Model 4 of Table 4 supports our hypothesis that the reviews in the spotlight position have additional impact on product sales. We also note that, in this model, the coefficient on overall star rating is not significant; suggesting the most of the consumer response is being explained by the combination of spotlight reviews and other reviews with high helpful votes. Also, the degree of impact of the average rating of spotlight reviews is stronger than the impact of the average rating of reviews that have high helpful votes. This result also implies that customers rely more on spotlight reviews than on other overall reviews.¹¹

Finally, we test Hypothesis 5 that customer reviews with a high proportion of helpful votes have a larger impact on less popular books than on more popular ones, we flag our samples into books that had a sales rank of less than 100,000 and books that had a sales rank of greater than 100,000 on the release date. We choose 100,000 because this is the number of unique titles normally carried by Barnes and Noble superstores (Brynjolfsson et al. 2003).¹² We flag those books using a dummy variable indicating a non-popular (high rank) book. We also add interaction terms between high rank dummy variable and other variables.

¹¹ We note, however, that this result should be interpreted with caution because of the relatively high correlation between spotlight reviews and reviews with a high proportion of helpful votes.

¹² We note that our qualitative result is not sensitive to the choice of rank cutoff.

Our results (Table 5) are consistent with Hypothesis 2, that reviews with high proportion of helpful votes have an additional marginal impact on sales. Moreover, the coefficient on the interaction term between average star rating of reviews with high helpful votes and high rank dummy variable is significant and has the same sign as average star rating of reviews with high helpful votes. This result means that the impact of average star rating of reviews with high helpful votes on sales of non popular books is 0.4048 more than that of popular books.

Table 4: Statistical Analysis				
Term	Model 1	Model 2	Model 3	Model 4
Ln (Amazon Price)	0.5502**	0.6069**	0.7437**	1.2232
	(0.0826)	(0.0890)	(0.1134)	(0.6472)
Ln (Days since release)	0.0696**	0.0504**	0.0952**	-0.0116
	(0.0102)	(0.0113)	(0.0169)	(0.0191)
Ship within 24 hours	-0.5150**	-0.5315**	-0.8551**	-0.4336**
	(0.0517)	(0.0568)	(0.1225)	(0.1399)
Ln (No. of reviews)	0.6404**	0.6833**	0.7955**	1.1048**
	(0.0280)	(0.0344)	(0.0637)	(0.0719)
Average star rating of all reviews	-0.3063**	-0.3751**	-0.4382**	-0.1672
	(0.0419)	(0.0575)	(0.1101)	(0.1482)
Average star rating of reviews that have more than 80% helpful vote		-0.5000**	-0.4236**	-0.2979**
		(0.0522)	(0.0681)	(0.0711)
Average star rating of reviews by top 1000 reviewers			0.0801	
			(0.0732)	
Average star rating of spotlight reviews				-0.7216**
				(0.0705)
N	13,783	10,798	5,830	4,750
R-Squared	0.0883	0.0982	0.1027	0.1069

Notes. Dependent variable is $\ln(\text{rank}) - \ln(\text{rank at the first day the book was released})$. All models run with book-type fixed effects. Standard errors are in the parentheses. * $p < 0.05$; ** $p < 0.01$.

Model 5 also shows that there are no statistically significant differences between the impact of price and shipping to sales of popular books and non popular books. The coefficient of the number of days since the book was released is positive, suggesting that sales decline over time, consistent with results from previous models. The coefficient on the interaction term between the number of days since the book was released and the high rank dummy variable is negative and significant, implying that sales of non-popular books are less sensitive to the passage of time. Specially, adding the interaction term's coefficient to the coefficient on the number of days since the book was released yields a small negative coefficient (-0.08), which means that sales for non-popular books are increased over time but the rate of increase will be very small or almost zero. This result gives the signal that the non popular books have the same rate of sales over a period of time. The coefficient on the number of reviews has the same sign as results from the previous models and the interaction term is significant. This coefficient implies that the marginal impact of an additional review declines with the number of reviews. Lastly, the coefficient of overall average star rating has the same negative sign as the results from previous models, but the interaction term of overall average star rating and high rank dummy is positive and significant. This result means that the impact of the overall average star rating on sales of non-popular books is less than that of popular books. On the other hand, we find that the impact of average star rating of high helpful votes are significantly stronger on non-popular books than on popular books.

Table 5: Additional Statistical Analysis	
Term	Model 5
Ln (Amazon Price)	0.8524**
	(0.1634)
Ln (Amazon Price) x Highrank	-0.2923
	(0.1943)
Ln (Days since release)	0.4767**
	(0.0234)
Ln (Days since release) x Highrank	-0.5569**
	(0.0265)
Ship within 24 hours	-0.5338*
	(0.2588)
Ship within 24 hours x Highrank	0.1784
	(0.2651)
Ln (No. of reviews)	0.3580**
	(0.0599)
Ln (No. of reviews) x Highrank	0.2376**
	(0.0734)
Overall average star rating	-0.5241**
	(0.1032)
Overall average star rating x Highrank	0.3127*
	(0.1233)
Average star rating of reviews with high helpful vote	-0.3014**
	(0.0714)
Average star rating of reviews with high helpful vote x Highrank	-0.4048**
	(0.1019)
N	10,798
R-Squared	0.1433

Notes. Dependent variable is $\ln(\text{rank}) - \ln(\text{rank at the first day the book was released})$. All models run with book-type fixed effects. Standard errors are in the parentheses. * $p < 0.05$; ** $p < 0.01$.

Discussion and Conclusion

Online feedback mechanisms and virtual communities have become increasingly important to consumers' decision making in online markets. However, most of the extant literature on online feedback mechanisms focuses on product or seller reviews using aggregate measures of quality and reputation. Less is known in the literature about how the quality of individual reviews and the reputation of individual reviewers influence the community's perception of the validity of the opinions expressed in the review. While the social psychology literature has shown that credibility influences the impact of a persuasive message, where credibility can be based either on the reputation of the author or the content of the message, this theory has not been examined empirically in the context of online markets (Guadagno et al. 2003). In extending these two streams of the literature, the goal of this paper is to examine the

micro-level impact of reviews — specifically the quality of online product reviews and the reputation of the reviewers — on sales.

Our results show that while higher book ratings are associated with higher book sales, higher quality reviews (i.e. reviews with the high proportion of helpful votes) strengthen this impact by creating additional sales. This result suggests that consumers may consider quality reviews more important in making purchase decisions. Furthermore, these reviews affect non-popular books more than popular books. We also find evidence that “Spotlight” reviews have a stronger effect on book sales than overall reviews do. However, contrary to our expectations, we find no evidence that the reputation of reviewers (i.e. top reviewers) is an important factor in consumers’ purchase decisions. We speculate that this may be due to the fact that top 1,000 reviewers must review so many products that they aren’t likely to have the requisite expertise in any product to make a significant impact on customer purchase decisions.

This research makes two unique contributions to the literature. First, this study uses a new dataset from a working online market to unpack the mechanisms that drive people to trust and respond to product reviews. To the best of our knowledge, this has not been studied empirically in the literature. Second, this research allows us to understand what form of micro-level dynamics of community interactions may be valuable in signaling quality — in addition to the aggregate-level summary quality scores.

For online retailers, our research suggests that community peer-rating systems provide an important signal of trust and can facilitate commerce. As noted above, the fact that the content of reviews matters to consumer purchase decisions should strengthen the reliability of online review systems by making it harder for self interested parties to manipulate the ratings. Ratings that provide a simple 5-star (or 1-star) review, while having equal weight in the overall average star rating listed on Amazon’s site, do not have as much influence on consumer response as more detailed reviews that have been rated as “helpful” by other members of the community. Even though the complete elimination of strategic manipulation of online reviews is difficult, Dellarocas (2006) suggests that if the unit cost of manipulation is high and the fraction of consumers who submit feedback grows, the level of manipulation will decrease because firms will be better off when consumers expect them to manipulate less. In our study, social validation (the number of helpful votes) can serve as an anti-manipulation tool that can increase manipulation costs. Manipulation costs might increase because self interested parties would need to invest more time in registering new user identities and ensuring that their written reviews were able to garner helpful votes. Likewise, aggregating the number of people who submit reviews and the number of people who vote if reviews are helpful can increase the fraction of consumers who submit feedback. As a result, the review voting systems can reduce the benefits to self interested parties from manipulating product reviews.

Our research also shows that reviews have the more impact on less popular books than on other titles. An implication of this finding for the publishing industry is that online review systems may play an important role in the development of “long tail” markets. Recent papers in the academic literature (Brynjolfsson, Hu, and Smith 2003) and business (Brynjolfsson, Hu, and Smith 2006)) and popular press (e.g., Anderson 2004; Anderson 2006) have discussed the impact of the increased product variety available in online markets on consumer surplus and industry structure. The Internet allows retailers to “stock” far more products than what would be possible in a typical brick-and-mortar environment. In the case of books, Internet booksellers can stock nearly all of the approximately 3 million books in print and numerous out-of-print titles while a typical brick-and-mortar bookstore can only stock 40,000 to 100,000 titles (Brynjolfsson, Hu, and Smith 2003). However, in the absence of reliable product information, it may be difficult to credibly signal the quality of these products to consumers. Online product review systems may be able to serve this function and play an important role in extending long tail markets — which can have important spillover effects for authors and publishers (Brynjolfsson, Hu, and Smith 2006).

Looking forward, the advertising literature (e.g., Akerberg 2003) notes that “informative/prestige” and “image” are two general effects of advertising. Informative advertising affects the consumer learning model while prestige or image advertising affect consumers’ utility functions. To isolate these two affects, it would be useful for future research to analyze the content of each review to see if it gives information on the product or confers image and prestige to the product. Future research could also observe the causality between prominent reviewers and number of helpful votes, or the effect of reviews across book genres. It also would be interesting for future research, using experimental or behavioral methods, to analyze the impact of product review on the consumer’s micro-level choice processes, cognitive processes, and the dynamics and motivations of reviewer communities.

References

- Ackerberg, Daniel A. "Advertising, learning, and consumer choice in experience good markets: an empirical examination," *International Economic Review* 44(3), August 2003, pp. 1007 – 1040.
- Anderson, C. "The Long Tail," *Wired Magazine*, Issue 12.10, October 2004.
- Anderson, C. *The Long Tail: Why the Future of Business Is Selling Less of More*, Hyperion Press, New York, NY, 2006.
- Ba S., Pavlou, P. "Evidence of the effect of trust building technology in electronic markets: Price premiums and buyer behavior," *MIS Quarterly* 26(3), 2002, pp. 243-268.
- Bass, F.M. "A New Product Growth Model for Consumer Durables," *Management Science* 15, 1969.
- Brynjolfsson, E., Dick, A. A., Smith, M. D. "Search and Product Differentiation at an Internet Shopbot," Working Paper, Carnegie Mellon University, Pittsburgh, PA, 2004.
- Brynjolfsson, E., Hu, Y., Smith, M. D. "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety," *Management Science* 49(11), 2003, pp. 1580-1596.
- Brynjolfsson, E., Hu, Y., Smith, M. "From Niches to Riches: Anatomy of the Long Tail," *Sloan Management Review* 47(4), 2006, pp. 67-71.
- Chen, P-Y., Wu, S-Y. "The Impact of Online Recommendations and Consumer Feedback on Sales," *Working Paper*, Carnegie Mellon University, Pittsburgh, PA, 2005.
- Chevalier, J., Goolsbee, A. "Measuring prices and price competition online: Amazon.com and BarnesandNoble.com," *Quantitative Marketing and Economics* 1(2), 2003, pp. 203-222.
- Chevalier, J., Mayzlin, D. "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, Forthcoming (August), 2006.
- Cialdini, R.B. *Influence: Science and Practice (4th ed)*, HarperCollins, NY, 2000.
- Dellarocas, C. "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms," *Management Science* 49(10), 2003, pp. 1407-1424.
- Dellarocas, C. "Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms," *Management Science* 52(10), 2005, pp. 1577-1593.
- Despotovic, Z., Aberer, K. "Possibilities for Managing Trust in P2P Networks," *Swiss Federal Institute of Technology (EPFL) Technical Report IC/2004/84*, Lausanne, Switzerland, 2004.
- Duan, W., Gu, B., Whinston, A. B. "Do Online Reviews Matter? - An Empirical Investigation of Panel Data," *Working paper*. University of Texas at Austin, Austin, TX, 2005.
- Ghose, A., Smith, M., Telang, R. "Internet Exchanges for Used Books: An Empirical Analysis of Product Cannibalization and Welfare Impact," *Information Systems Research* 17(1), 2006, pp. 3-19.
- Ghose, A., Sundararajan, "A. Software Versioning and Quality Degradation? An Exploratory Study of the Evidence," *Working Paper*. New York University, NY, 2005
- Guadagno, R.E., Cialdini, R.B. "Online persuasion and compliance: Social influence on the Internet and beyond," In Y. Amichai-Hamburger (Ed.), *The social net: The social psychology of the Internet*. Oxford, UK: Oxford University Press, 2003.
- Matheson, K., Zanna, M. P. "Persuasion as a function of self-awareness in computer-mediated communication," *Social Behaviour* (4), 1989, pp.99-111.
- Paumgarten, N. "No. 1 Fan Dep't Acknowledged," *The New Yorker*, May 5, 2000.
- Poynter, D. "Publishing Poynters," April-June 2000. (Available at <http://parapub.com/getpage.cfm?file=newsletter/News0400.html>.), accessed February 26, 2003.

- Resnick, P., Zeckhauser, R. "Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system," *The Economics of the Internet and E-Commerce. Advances in Applied Microeconomics*, Vol. 11. 2002 JAI Press, Greenwich, CT.
- Resnick, P., Zeckhauser, R., Friedman, E., Kuwabara, K. "Reputation Systems," *Communications of the ACM* 43(12), 2000, pp. 45-48.
- Smith, M., Brynjolfsson, E. "Customer Decision Making at an Internet Shopbot: Brand Still Matters," *The Journal of Industrial Economics* 49(4), 2001, pp.541-558.
- Weingarten, G. 2001. Below the beltway, Washington Post, June 17, 2001.