Research Note:

All Reviews are Not Created Equal:

The Disaggregate Impact of Reviews and Reviewers at Amazon.com

Pei-yu Chen,[‡] Samita Dhanasobhon,[†] and Michael D. Smith[†]

{pychen, sdhanaso, mds}@andrew.cmu.edu

This Version: May 2008

[†] H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, Pittsburgh, PA, 15213. (Dhanasobhon: 412-268-1663; Smith: 412-268-5978)

[‡] Fox School of Business, Temple University, Philadelphia, PA, 19122. (Chen: 412-268-1291)

All Reviews are Not Created Equal:

The Disaggregate Impact of Reviews and Reviewers at Amazon.com

ABSTRACT

Online product review networks help to transmit information that customers can use to evaluate product quality. The prior literature has found that, in the aggregate, better product reviews lead to higher sales. However, product review networks increasingly include an explicit social component that allows consumers to evaluate individual reviews based on the status of the reviewer and on the "helpfulness" of the review to the community.

In this research, we extend this literature by analyzing the impact of reviews at a disaggregate level. We find that reviews that the community finds helpful have a stronger influence on consumers' purchase decisions than other reviews do. Moreover, these reviews have a stronger impact on less popular books than on more popular books, where consumers may be able to use other outside information sources to form an opinion of the product.

Overall, our results suggest that the micro-level dynamics of community interactions are valuable in signaling quality over-and-above aggregate-level scores. One implication of this result is that the micro-level dynamics of reputation communities make it harder for self-interested parties to manipulate reviews versus an environment where consumers only have aggregate quality measures.

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

1. 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 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). Applying these findings to online commerce, several papers have shown that online reviews influence the probability of purchase. For example, Resnick (2002) shows that seller reviews in eBay influence the probability of a sale, and Chevalier and Mayzlin (2006) find that product reviews at Amazon.com impact book sales.

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 itself 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' purchase decisions. 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 (see Figure 1: Amazon lists "x of y people found this review helpful"). 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 that are displayed on the product page (see Figure 2). These "spotlight" reviews are set apart and shown before other reviews on the product page, so they may have a relatively stronger effect on book sales than other reviews do.

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 review information has a stronger impact on niche titles than on more popular books. Finally, we find no evidence that the social status of reviewers, as displayed by Amazon.com, 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 that 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.

The remainder of this paper proceeds as follows. In Section 2, we review the prior literature as it relates to our research setting. In Section 3, we use this literature to develop our research framework and hypotheses. We present our data and model specifications in Sections 4 and 5 respectively. We present our empirical results in Section 6 and conclude in Section 7.

2. Literature Review and Research Framework

In digitally mediated markets, absent reputation systems, consumers can face high uncertainty about product quality, since there are in many cases fewer quality and trust cues available than what is possible in brick-and-mortar markets (Smith and Brynjolfsson 2000). 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). Moreover, when product quality cannot easily be verified, consumers can use the supplier's reputation as an indication of 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, self-interested behavior on the part of sellers may reduce the informativeness of product reviews. Jin and Kato (2006) note that consumers must be aware of the posted messages about product supplier/seller quality because the cost of switching identities can be low in online markets, allowing self-interested parties to strategically manipulate review systems. In order to address the potential for this sort of strategic manipulation in explicit feedback, distributed reputation mechanisms were adopted by some online sites. Distributed reputation mechanisms capture the community's assessment of the quality of individual reviews, and allow consumers to take these assessments into account in their purchase decision.[1]

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. They explain that consumer's tastes could be sufficiently heterogeneous that they do not use other consumers' opinion in their purchase decision, or that consumers many not find product ratings informative because most products receive relatively high ratings.

However, there remain important open questions about the micro-level impact of reviews online and in particular the impact of social cues in review communities on product sales. This paper aims to address these questions by developing empirical results that 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

[1] See Dellarocas (2005) and Despotovic and Aberer (2004).

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). We describe the theoretical basis for these analyses in more detail below.

3. Theoretical Framework

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, *a priori,* for the consumer to determine the real quality of the book. Instead the consumer must base her decision on observable signals of 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,

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

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 (which in our setting primarily includes product reviews at Amazon.com).

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: $r_1, r_2, r_3, \cdots, r_N$, 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 as follows:

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

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 this as an empirical question to be determined by our data.

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

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

where $\bar{q}_R$ is the derived quality measure weighted by reviewer reputation, and the first part, $\bar{q}_A$, is an aggregate measure of quality without taking into account reviewer reputation or content quality (see equation 5), while the latter part captures the impact of reviewer reputation beyond the aggregate measure.

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 as follows:

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

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 (as 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.,

$$\overline{q}_A = \overline{r} = \frac{\sum_{i=1}^{N} r_i}{N} \tag{5}$$

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

The main contribution of this paper is to analyze the importance of content quality and reviewer reputation, and to empirically measure the role of content quality and reviewer reputation in influencing consumer decisions after controlling for the average ratings.

To this end, given equations (1) and (5), we have:

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

When any $w_i \delta_i, \forall_i \in \{1,2,\ldots,N\}$ deviates from $1/N$, it indicates that consumers take into account the content quality and reviewer reputation when evaluating products based on online reviews. We can further test the individual impact of content quality and reviewer reputation by using equations (3) and (4) to consumer decision-making. For example, by examine the relationship between $\overline{q}_R = \overline{q}_A + \sum_{i=1}^{N} r_i \left(w_i - \frac{1}{N} \right)$ and the consumer's purchase decision. One can show that if any $w_i, \forall_i \in \{1,2,\ldots,N\}$ deviates from $1/N$, it follows that reviewer reputation does matter in consumer decision-making. Moreover, by examining the relationship between $\overline{q}_C = \overline{q}_A + \sum_{i=1}^{N} r_i \left(\delta_i - \frac{1}{N} \right)$ and the consumer's purchase decision, one can show that if any $\delta_i, \forall_i \in \{1,2,\ldots,N\}$ 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.

We now apply this framework to develop the theoretical hypotheses we test in this study.

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

This hypothesis was studied in Chevalier and Mayzlin (2006), who found that higher product ratings are associated with higher sales. We include it here to primarily to the frame our main results below concerning the disaggregate impact of product reviews.

From a theoretical perspective, higher product ratings ($\bar{q}_A$) convey two signals: they may indicate a high quality of the book, and they also imply that the general public likes 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 that has not.

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

> **H2a: High star ratings from reviews with high proportion of helpful votes result in an increase in sales.**
>
> **H2b: Low star ratings from reviews with high proportion of helpful votes result in a decrease in sales.**

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 after they have purchased the product. We suggest that the provision of the quality of review content can be used as a sanctioning device to alleviate the 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. Specifically, we empirically test 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 their opinion of the helpfulness of the review by voting "Helpful" or "Not helpful."[2] 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.

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. Amazon determines the ranking of reviewers 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 "helpful" votes. A reviewer who reviews only best-selling items will receive more votes than a reviewer who only takes more obscure items into consideration, and Amazon takes these

[2] It is important to note that "helpfulness" in this context should be understood as "help in making a purchase decision." That is either positive and negative evaluations of the product could be "helpful" to the consumer.

differences into account 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 prominent reviewers as any reviewer identified as a "top 1,000" reviewer. This serves as a useful cutoff as Amazon shows a rank badge next to any reviewer listed among the top 1,000 reviewers (Figure 1).[3]

The importance of these reputation cues derives directly from the literature. 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 computer-mediated decision groups than in face-to-face interaction, Guegen and Jacob (2002) showed that status and expertise create higher compliance, especially when messages come from a high-status member. Guadagno and Cialdini (2003) summarized these results as follows: "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) have higher influence over consumer decisions than other reviewers do.

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

In its online review listings, Amazon.com identifies the two "best" reviews, and places these reviews before the others, highlighting them as "Spotlight" reviews. These spotlight reviews are updated as additional consumers cast helpful votes on individual reviews. Thus, the most recent

[3] Our results are not sensitive to this assumption, and would remain essentially unchanged if we used a lower cutoff to identify "top" reviews.

voting history determines which reviews are identified as spotlight reviews. 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. 2004) and that the order information is displayed to the consumer has a disproportionately strong impact on their behavior (Smith and Brynjolfsson 2001). Both of these results should hold 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 a consumers' prior beliefs of product quality on their decision-making. When consumer has a more confident prior, α is likely to be higher, and $\overline{q}_M$ will therefore be less important in the consumers' decision making. One proxy for the strength of a consumer's prior is the popularity of the book. 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 evaluating these books may come to Amazon with a stronger prior belief about their quality. For less popular books, particularly newly released titles, consumers may 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 they do on more popular books.

4. Data

Our data are collected from publicly available information on Amazon.com's book marketplace, 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 new books (20 titles per week x 28 weeks) released over a 195-day period from November 11, 2005 to May 25, 2006.[4] We focus on newly released titles because consumer opinions are less well formed for these products, making product reviews potentially more important for consumer purchase decisions; and because there are significant changes in the number of reviews for these titles (initially zero, increasing over time), providing an additional source of variation in our data.

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

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.[5] For this sample, we extract

[4] Experimentation with different temporal sub-samples shows that our results are not sensitive to seasonality.
[5] 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.

generic information for 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. In addition, for each book in our sample we collect daily information from Amazon.com on the price, sales rank, and the time until the book would ship.

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.[6] 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:

$$Quantity = \beta_1 Rank^{B_2} \tag{7}$$

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 β_2 = -0.954.[7] This estimate is in the

[6] This technique has also been applied in a variety of other studies, including Chevalier and Mayzlin (2006), Ghose, Smith, and Telang (2006), and Ghose and Sundararajan (2005)
[7] 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 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.

range of coefficient values reported by other studies in the literature (albeit at the high end of this range).[8]

Table 2: Summary Statistics

Variable	*Obs.*	*Mean*	*Std. Dev.*	*Min*	*Max*
No. of days since release	14,187	64.66	42.22	1	188
Sales rank	14,187	495,210	744,254.6	4	3,934,116
Amazon Price	14,187	19.01	20.75	3.99	199
Number of reviews	14,187	9.567562	16.27	1	170
Average star rating of all reviews	14,187	4.40	0.72	1	5
Average star rating of reviews that have more than 80% helpful vote	11,095	4.55	0.72	1	5
Average star rating of reviews by top 1000 reviewers	6,800	4.59	0.55	2	5

Table 3: Correlation Matrix

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sales rank (1)	1							
Amazon Price (2)	0.2937	1						
No. of days since release (3)	0.1196	0.046	1					
Shipping (4)	-0.4033	-0.2696	0.1268	1				
No. of reviews (5)	-0.3365	-0.0783	0.173	0.1508	1			
Average star rating of all reviews (6)	0.185	0.0466	-0.1339	-0.1049	-0.4489	1		
Average star rating of reviews with more than 80% helpful vote (7)	-0.1143	-0.3065	-0.0002	0.0952	-0.0278	0.0899	1	
Average star rating of reviews by top 1000 reviewers (8)	0.0002	0.0134	-0.0388	0.1551	-0.1255	0.2286	0.2949	1
Average star rating of spotlight reviews (9)	0.0526	-0.1293	0.0413	-0.0544	-0.2888	0.3484	0.6975	0.4293

Finally, we collect daily information regarding the reviews posted at Amazon for each book in our sample. For each book review, we collect the review's posting date, the full text of the

[8] 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.

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),[9] the number and proportion of helpful votes (see Figure 1), and whether the review was highlighted as a "spotlight" review (see Figure 2).[10] Table 2 provides summary statistics for our data and Table 3 provides a correlation matrix for these variables.

5. Methodology

Our empirical approach is based on Chevalier and Mayzlin (2006), but with estimating the model only for Amazon.com data. Specifically, to study the impact of reviews and the quality of reviews on sales, we consider the following model:

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

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 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.

Within this model, we use different measures of review activity (R) to fit the model. In our base model we use the difference between the number of reviews on the first day after release and that

[9] 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.
[10] 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.

on day *t* and the difference between average star rating across all reviews on the first day and that on day *t*. We extend this to measure of 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 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.

We run these models using book level fixed effect. 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), thus one can consider the supply of books to be fixed and exogenous at the time the retailers sets prices. Because of this, we follow Ghose et al. (2006) and do not consider price and quantity as jointly determined in our models. We note, however, 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.

6. Results

We now fit these empirical models to our data in Tables 4 and 5.[11] As noted earlier, a positive coefficient means a variable of interest is positively associated with sales rank and but negatively associated with sales. In these results, our control variables are consistent across specifications and have the expected signs. Notably, the coefficient on price is positive and significant across

[11] We note that the R^2 values in our models are comparable to Chevalier and Mayzlin (2006) reported R^2 values models.

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).[12]

With respect to our review variables of interest, as expected we find that higher overall star ratings have a positive impact of sales (negative impact on sales rank), consistent with Chevalier and Mayzlin's (2006) results. Extending these results, in Model 2-4 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 suggests that reviews which are identified by the community as "helpful" are more influential on sales than are average star ratings (which take into account every review posted on the site and which weigh each review equally). Using model 2 as a reference point, we see that while average star ratings are associated with higher sales (-0.375), average star ratings among reviews with more than 80% helpful votes are associated with an additional marginal increase in sales (-0.5). Overall, these results show that, consistent with hypothesis 2, quality reviews (i.e., reviews with high helpful votes) are more influential on sales than other reviews are.

In Model 3 of Table 4, we include a variable for the average star rating among reviewers who are identified in 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

[12] 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.

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.

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 an additional impact on product sales, above and beyond aggregate star ratings. 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 may rely more on spotlight reviews than on other overall reviews.[13]

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. To do this, we include a dummy variable in our data, which identifies books that have a sales rank of greater than 100,000 on the release date. We choose 100,000 as the cutoff rank because this is the number of unique titles normally carried by Barnes and Noble superstores (Brynjolfsson et al. 2003) and has been used in the past to identify "long tail" books; however our results are not sensitive to this cutoff.

[13] We note, however, that this result should be interpreted with caution because of the relatively high correlation in Table 3 between spotlight reviews and reviews with a high proportion of helpful votes.

We then add interaction terms between high rank dummy variable and other variables to see whether there is any systematic difference in response to ratings between less popular and more popular books.

Table 4: Statistical Analysis

Term	Model 1	Model 2	Model 3	Model 4
Ln (Amazon Price)	0.5551**	0.6095**	0.7437**	1.2232
	(0.0798)	(0.0857)	(0.1134)	(0.6472)
Ln (Days since release)	0.0713**	0.0533**	0.0952**	-0.0116
	(0.0100)	(0.0112)	(0.0169)	(0.0191)
Ship within 24 hours	-0.4912**	-0.5028**	-0.8551**	-0.4336**
	(0.0503)	(0.0551)	(0.1225)	(0.1399)
Ln (No. of reviews)	0.6371**	0.6786**	0.7955**	1.1048**
	(0.0277)	(0.0342)	(0.0637)	(0.0719)
Average star rating of all reviews	-0.2914**	-0.3715**	-0.4382**	-0.1672
	(0.0378)	(0.0572)	(0.1101)	(0.1482)
Average star rating of reviews that have more than 80% helpful vote		-0.4948**	-0.4236**	-0.2979**
		(0.0518)	(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	14,187	11,095	5,831	4,750
R-Squared	0.0883	0.0963	0.1027	0.1069

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

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. We discuss these results in more detail below.

Table 5: Additional Statistical Analysis

Term	Model 5
Ln (Amazon Price)	0.8524**
	(0.1628)
Ln (Amazon Price) x Highrank	-0.2932
	(0.1910)
Ln (Days since release)	0.4767**
	(0.0233)
Ln (Days since release) x Highrank	-0.5508**
	(0.0264)
Ship within 24 hours	-0.5338*
	(0.2579)
Ship within 24 hours x Highrank	0.2022
	(0.2638)
Ln (No. of reviews)	0.3580**
	(0.0596)
Ln (No. of reviews) x Highrank	0.2284*
	(0.0731)
Overall average star rating	-0.5241**
	(0.1028)
Overall average star rating x Highrank	0.3176*
	(0.1228)
Average star rating of reviews with high helpful vote	-0.3014**
	(0.0712)
Average star rating of reviews with high helpful vote x Highrank	-0.3940**
	(0.1014)
N	11,095
R-Squared	0.1402

Notes. Dependent variable is ln(rank)- ln(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.

7. Discussion and Conclusion

Online feedback mechanisms and virtual communities have become increasingly important to consumers' ability to evaluate products 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, Guadagno et al. (2003) note that this theory has not been examined empirically in the context of online markets (Guadagno et al. 2003). Thus, the goal of this paper is to extend these two literatures by examining the micro-level impact of reviews on sales, with a particular focus on how quality of online product reviews and the reputation of the reviewers influence consumer's decisions.

Our results show that while higher ratings are associated with higher book sales, higher quality reviews (i.e. reviews with the high proportion of helpful votes) have a stronger impact on consumer purchase decisions than other reviews do. This result suggests that consumers may consider higher quality reviews more important in making purchase decisions (over and above the aggregate "star rating" for the product). We also find evidence that, *ceteris paribus*, "Spotlight" reviews have a stronger effect on book sales than overall reviews do — which is consistent with the hypothesis that consumers may economize on costly search by focusing their attention on a few highlighted reviews. 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 sufficiently many products that they aren't likely to have the requisite expertise in any specific 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 online "rate the reviewer" systems provide an additional signal of trust to consumers beyond aggregate quality scores. As noted above, the fact that the content of individual 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 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 (Brynjolfsson, Hu, and Smith 2003) and business literatures (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 serve this function and thus may play an important role in extending long tail markets — with potentially important spillover effects for authors and publishers.

Looking forward, the advertising literature (e.g., Ackerberg 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 Marking 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.

Dubrovsky, V.J., Kiesler, S., Sethna, B.N., "The Equalization Phenomenon: Status Effects in Computer-Mediated and Face-to-Face Decision-Making Groups," *Human-Computer Interaction* 6(2), June, 1991, pp. 119 - 146.

Etzion, H., Awad, N. "Pump up The Volume? Examining the Relationship between Number of Online Reviews and Sales. Is More Necessarily Better?," *Proceeding International Conference on Information Systems 2007.*

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.

Guéguen, N., Jacob, C., "Solicitation by e-mail and solicitor's status: a field study of social influence on the web,"*Cyberpsychology Behavior* 5(4), 2002, pp. 377-83.

Hu, N., Pavlou, P. A., Zhang, J. "Can Online Reviews Reveal a Product's True Quality? Empirical Findings and Analytical Modeling of Online Word-of-Mouth Communication," *EC,* June 11–15, 2006.

Jin, G. Z., Kato, A. "Price, quality, and reputation: evidence from an online field experiment," *The RAND Journal of Economics* 37(4), 2006, pp. 983 - 1005.

Li, X., Hitt L. M. "Self Selection and Information Role of Online Product Reviews," Information Systems Research Special Issue: The Interplay Between Digital and Social Networks.

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.

Stiglitz, J.E. "Imperfect Information in the Product Market," *Handbook of Industrial Organization Vol. 1*, *Elsevier Science Publishers*, 1989, pp. 769-847.

Weingarten, G. 2001. Below the beltway, Washington Post, June 17, 2001

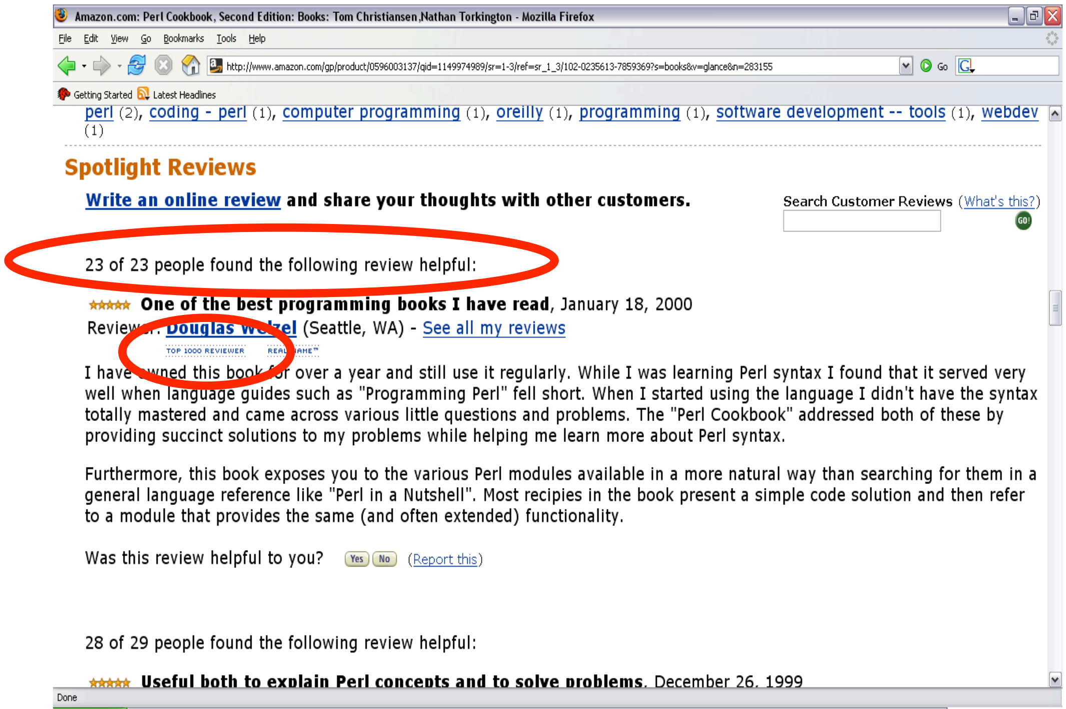

Figure 1: Number of Helpful Votes and Top Reviewer Badge at Amazon.com

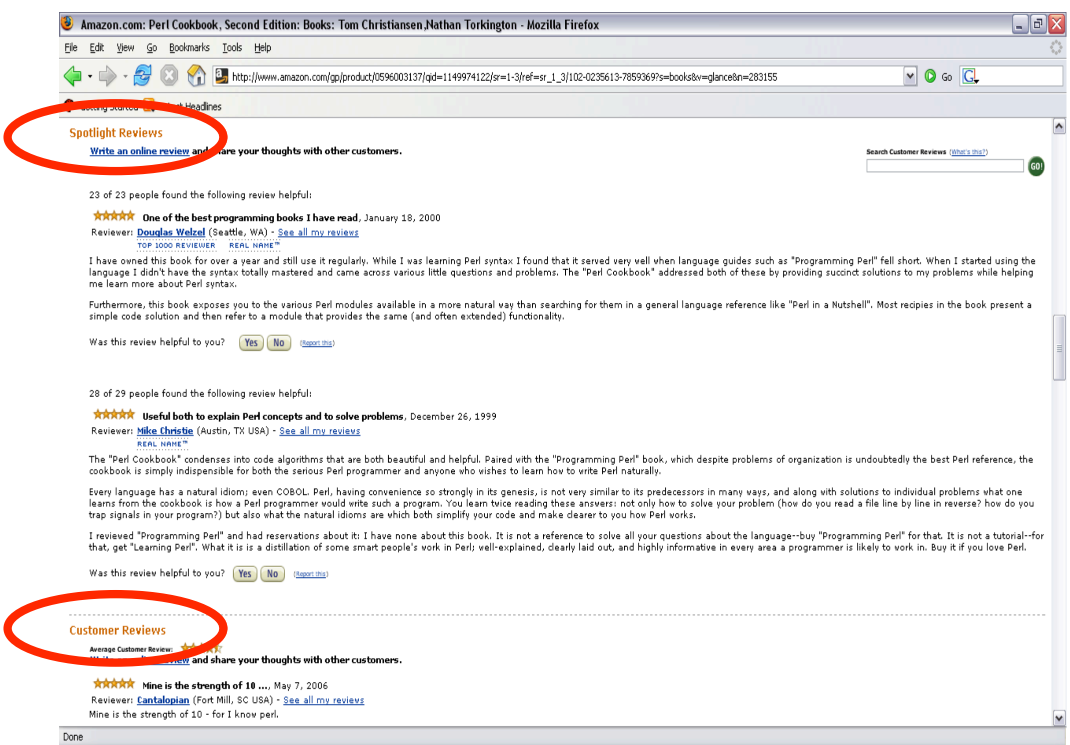

Figure 2: Sample Spotlight Reviews and Customer Reviews at Amazon.com