

Yale SCHOOL *of* MANAGEMENT

Working Paper Series ES & MK Economics & Marketing

“The Effect of Word of Mouth on Sales: Online Book Reviews”

Judy Chevalier, *Yale School of Management, National Bureau of
Economic Research (NBER)*

Dina Mayzlin, *Yale School of Management*

Working Paper ES # 28 & MK #15

This paper can be downloaded without charge from the
Social Science Research Network Electronic Paper Collection:

<http://ssrn.com/abstract=432481>

THE EFFECT OF WORD OF MOUTH ON SALES: ONLINE BOOK REVIEWS *

Judith Chevalier
judith.chevalier@yale.edu

Dina Mayzlin
dina.mayzlin@yale.edu

Yale School of Management
135 Prospect Street
P.O. Box 208200
New Haven, CT 06520-8200

August 6, 2003

* This is a preliminary draft. Please do not quote without the authors' permission. We would like to thank participants at Marketing Science Conference in U. of Maryland, and Sharon Oster for helpful comments. Both authors contributed equally and their names are listed in alphabetical order.

THE EFFECT OF WORD OF MOUTH ON SALES: ONLINE BOOK REVIEWS

The creation of online consumer communities to provide product reviews and advice has been touted as an important, albeit somewhat expensive component of Internet retail strategies. In this paper, we characterize reviewer behavior at two popular Internet sites and examine the effect of consumer reviews on firms' sales. We use publicly available data from the two leading online booksellers, Amazon.com and BarnesandNoble.com, to construct measures of each firm's sales of individual books. We also gather extensive consumer review data at the two sites. First, we characterize the reviewer behavior on the two sites such as the distribution of the number of ratings, the valence and length of ratings, as well as ratings across different subject categories. Second, we measure the effect of individual reviews on the relative shares of books across the two sites. We argue that our methodology of comparing the sales and reviews of a *given* book *across* Internet retailers allows us to improve on the existing literature by better capturing a causal relationship between word of mouth (reviews) and sales since we are able to difference out factors that affect the sales and word of mouth of both retailers, such as the book's quality. We examine the incremental sales effects of having reviews for a particular book versus not having reviews and also the differential sales effects of positive and negative reviews. Our large database of books also allows us to control for other important confounding factors such as differences across the sites in prices and shipping times.

Keywords: advertising, word-of-mouth, source credibility, Internet marketing

1. Introduction

Online user reviews have become an important source of information to consumers, substituting and complementing other forms of word of mouth communication about the quality of various products. Consequently, many managers believe that a Web site needs to provide community content in order to build brand loyalty. (See, for example, McWilliams (2000) or Fingar, Kumar, and Sharma (2000)). Despite this widespread belief, to our knowledge, there is no literature documenting that community content plays any role in consumer decision-making. Such a finding, it seems, is a necessary prerequisite for content provision to be a profitable strategy.

There are many reasonable arguments as to why making investments in providing such content could potentially be a poor strategy. First, it is not clear why users would bother to take the time to provide reviews for which they are not in any way compensated. Second, even if user reviews are provided, rival retailers can free ride on them; there is nothing to stop a consumer from utilizing the information provided by one website to inform purchases made elsewhere. Third, by providing user reviews, a site cedes control over the information displayed; unfavorable reviews created by either legitimate users or by biased interested parties may depress sales.¹ Similarly, since authors and publishers can freely proliferate favorable reviews for their own books, positive reviews may not be credible and may not function to stimulate sales. Last, online user reviews may not be useful, and may not stimulate sales due to the sample selection bias that is inherent in an amateur review process. That is, a consumer only chooses to read a book or watch a movie if she perceives

¹ See Mayzlin (2003) for a theoretical treatment of recommendation systems where firms can anonymously post reviews.

that there is a high probability that she will enjoy the experience. In the presence of consumer heterogeneity, this implies that the pool of reviewers will have a positive bias in their evaluation compared to the general population. Thus, positive reviews may simply be discounted by potential buyers.²

In this study, we characterize patterns of reviewer behavior, and examine the effect of consumer reviews on firms' sales patterns. In particular, we use publicly available data from the two leading online booksellers, Amazon.com and BarnesandNoble.com, to construct measures of each firm's sales of individual books. Both BarnesandNoble.com and Amazon.com provide customer reviews. However, Amazon.com's investments in "collaborative" consumer content have been more extensive and much-imitated by other Internet retailers. For example, Amazon.com has many strategies in place to try to elicit more and better-quality consumer content. By focusing on the differences between the two sites' sales of the same books, we examine the relationship between the customer reviews at each site and firm sales, controlling for other drivers of book sales.

Our work contributes to the broader literature on the link between customer word of mouth and sales, which has been demonstrated in several studies. For example, researchers have used word of mouth to explain the adoption of high-yield varieties of seeds among farmers (Foster and Rosenzweig (1995)), the adoption of tetracycline among physicians (Coleman (1966)), as well as evolution of the ratings of new TV shows (Godes and Mayzlin (2003)).

² In a very different context, Resnick and Zeckhauser (2002) find that 99% of the feedback ratings on ebay.com are positive.

However, these studies have an important limitation in that they do not determine the direction of causality between word of mouth and product sales. Theoretically, this direction is not clear: word of mouth may be either the driver or a leading early indicator of total sales. For example, in herding models such as Banerjee (1992) and Bikhchandani et al. (1991) relatively small differences in signals received by the customers who initially sample the product may have lasting long-range consequences on the success or failure of a product – the early trials drive total sales. In their model, then, word of mouth is an important driver of sales. Alternatively, the initial word of mouth may be a (noisy) signal of over-all performance since a product’s quality is revealed with time - word of mouth is an early measure of a product’s success but not necessarily its driver.³ In support of this view, Van den Bulte and Lilien (2001) re-analyze Coleman’s data to demonstrate that word of mouth influence on tetracycline adoption was over-estimated in the original study due to a lack of control for the marketing efforts of the drug companies.

In this study, we are better positioned to credibly establish the causality between word of mouth and sales by comparing the sales of a given book across the two booksellers. Hence, we are able to measure the effect of reviews on the sales of a book at each site, holding book quality constant. For example, suppose that a publisher engages in a heavy promotional off-line campaign for a particular book. This advertising campaign might both stimulate sales of the book and stimulate customers to post positive reviews online. The “traditional” methodologies for examining the affect of customer “word of mouth” would be to examine the sales of this book and the online reviews either through time, or in comparison to other books. These methodologies would suggest a positive relationship between customer

³ For example, Eliashberg and Shugan (1997) show that critical acclaim seems to serve as an early indicator of a movie’s over-all box office success.

reviews and book sales, but this relationship would not be causal. Both the sales and the reviews are really just the outcome of the publisher's promotional campaign.

In our methodology, we examine the relationship between market shares and customer reviews for a *given* book *across* the two sites. By focusing on the *differences* between the market share of the book at the two sites, we are able to control for shocks to word of mouth and to sales that are common to both booksellers and, instead, focus on the idiosyncratic shocks alone. Consider a book that is generally well-reviewed and well liked. If a cranky consumer posts a negative review of that book on Amazon, but doesn't post that review on Barnes and Noble.com, will the market share of the book at Amazon fall relative to the market share of the book at Barnesandnoble.com? This "ideal experiment" is the basis of our empirical strategy. Of course, data limitations force our analysis to differ somewhat from the ideal experiment, as we discuss later. However, we observe the same books, their customer reviews, and a proxy for each book's market share at each site. Our large database of books also allows us to control for other important factors that might affect the relative market share of a particular book across sites, such as differences across the sites in the price of the book or differences in the speed with which the book has been promised to be shipped. Furthermore, in order to partially rule out the hypothesis that the differences in word of mouth across sites are driven by unobservable underlying differences in the two populations, we show that the two sites are very similar in terms of customer preferences and reviewer behavior across broad categories of books in our sample.

Note that our focus on the differential market share of a given book across the two sites potentially greatly under-estimates the effect of word of mouth on sales. Since we estimate

the effect of online reviews on the market share of a book at the site on which this review appears, we ignore the possibility of free-riding. For example, Barnes & Noble.com customers could read Amazon reviews, or, similarly, Amazon reviews could affect offline sales. In fact, the success of a recently released best-seller “DaVinci Code” was attributed partly to an endorsement by a prolific Amazon reviewer: Francis McInerney.⁴ In this sense, our measure of the possible effect of word of mouth on sales is a conservative one.

Our user review data contains a star rating provided by the reviewer as well as a text description. In this paper, we focus our analysis on the star ratings. In fact, operationally, the star ratings provide an excellent opportunity to measure the valence of comments without analyzing the comments themselves, a very difficult task as demonstrated in Godes and Mayzlin (2003). We examine the incremental sales effects of having reviews for a particular book versus not having reviews and also the differential sales effects of positive and negative reviews.

The rest of the paper is organized as following. In Section 2, we describe the data. In Section 3, we describe the methodology and present results on the distribution of reviews and sales across sites, providing further insight into the reviewing process. In Section 4, we present our empirical analysis of the effect of “word of mouth” reviews on product sales, and, lastly, we conclude.

⁴ Paumgarten, N. “No. 1 Fan Dep’t Acknowledged,” www.newyorker.com, Issue of 2003-05-05, posted 2003-04-28.

2. Data

Our data consists of individual book characteristics and user review data that were collected from the public Web sites of Amazon.com and BarnesandNoble.com. The goal was to generate a representative sample of sites' sales. Since we do not have access to this proprietary data, we approximate a random sample of sales in the following way. First, we collect a random sample of books released. In order to maximize the probability that a book would be available on both Amazon and Barnes and Noble, we focus on a set of relatively recent books: titles that were released in the last five years. One shortcoming of a random sample of published books is that it overweighs books that have very few sales. Thus, in addition, we also extract a sample that consists of books that appeared at least once on a bestseller list. Hence, the sample was generated from two sources:

- 1) A random sample of books selected from a catalog "Global Books in Print" that were published in 1998-2002. (See Appendix for description of algorithm to generate the sample).
- 2) Publisher's Weekly bestseller lists: titles that appeared in the lists from 1/14/1991 to 11/11/2002.

Since a given book can be released in many different formats (such as hardback, paperback, etc), we use data from Bowker's Global Books in Print.com to obtain a listing of all possible English-language format releases of a given book. We discarded digital and audio format releases. Fortunately, each title-format combination has a unique International Standard Book Number (ISBN), assigned under the auspices of the International ISBN Agency in Berlin. Though it may not be apparent to the casual user, both Amazon.com and Barnes

and Noble.com use the ISBN numbers to organize the cataloguing of books on their web sites.

Over a two-day period in May of 2003, we searched the two Web sites to extract a body of data for each of the ISBN numbers in our sample.⁵ Our extraction included: the title, author, publisher, release date, and format type of the book. We also gathered information on the price charged for the book at each website, the promised time until the book would ship, and data for the most recent 500 reviews of the book posted on the website (we extracted the number of stars assigned, the date the review was posted, and the full text of the review). Most books have far fewer than 500 reviews, but for those with more than 500 reviews, we also extracted the total number of reviews posted, as well as the average number of stars assigned overall.

Lastly, both BarnesandNoble.com and Amazon.com provide a “sales rank” for each book on the site. These sales ranks reflect the total sales of that book at that site relative to the sales of other books at that site. Chevalier and Goolsbee (2003a) report that Amazon claims that for books in the top 10,000 ranks, the rankings are based on the last 24 hours and updated hourly. For books ranked 10,001-100,000, the ranks are updated once per day. For books ranked greater than 100,000, the sales ranks are updated once per month (Amazon, 2000). Based on this system then, books that have not been purchased in the past month would not be ranked. Many hundreds of thousands of books, however, have a rank but almost certainly have less than one sale per month. Italic (2001) claims that for these rarely purchased books, Amazon bases the rank the total sales since Amazon's inception. BN.com

⁵ The data on Amazon was collected on 5/6/2003 and the data on Barnes and Noble was collected on 5/7/2003.

claims to update all the rankings daily (BN.com, 2000).⁶ Thus, importantly for our purposes, with the exception of the books that have very high ranks on Amazon, the rankings represent a *current* snapshot of sales.

For this draft, we consider the set of ISBNs for which the most popular ISBN within a title is the same for both Amazon.com and BarnesandNoble.com. This creates a sample of 2505 ISBN codes. We leave aggregating ISBNs within a title to a future draft. Since we are not aggregating across books, we can use the sales ranks as is in our analysis, and discuss the impact of reviews on sales ranks directly. However, an extension of the methodology described in Chevalier and Goolsbee (2003a) and Schnapp and Allwine (2001) will allow us to also calibrate the sales rank relationships into total sales relationships. Any such calibration will be “back of the envelope”, but will give us an opportunity to understand very approximate magnitudes.

Schnapp and Allwine (2001) have proprietary data from a single publisher from May of 2001 relating that publisher’s sales at Amazon to that publisher’s sales ranks. They fit the sales-ranks relationship for a subsample of the publisher’s titles as:

$$\ln(\text{sales}^{\text{AMZN}}) = 9.61 - 0.78 \ln(\text{rank}^{\text{AMZN}})$$

While they do not provide R-squareds or other measures of fit, the scatterplots they supply suggest that the fit is very good and suggests no obvious objection to the underlying

⁶ Since BN.com provides rankings on tens of thousands of books that average far less than one sale per day, this statement cannot be completely accurate. They would not provide us any more detail in their ranking system (despite repeated requests).

distributional assumption. Since this dates from 2001, we scale up their estimates by 24%, the growth in Amazon.com's North American sales in the two years intervening between the time of our sample and the time of their sample. BN.com does not report data to publishers in a way that allows them to make such a comparison. We assume that the basic shape of the rank to sales relationship is the same at BN.com as it is at Amazon.com, but that it is scaled down to reflect the fact that BN.com's total sales equal about 15% of Amazon.com's North American sales.

We only include in our sample those books that are listed as "available" at both sites. Finally, we are forced to address the problem that BN.com only provides sales ranks for approximately 650,000 books and address the issue of "stale ranks" on Amazon. There are books at BN.com that are available for purchase but for which the rank is "too high" to be disclosed. Amazon does not censor their sales ranks and they appear to range upwards of one million. If we were to use as our sample all books with prices and ranks at both sites, our sample would contain a large number of books that are relatively popular at BN.com, and relatively unpopular at Amazon. However, books that are relatively popular at Amazon.com and relatively unpopular at BN.com would not appear in the sample, as they have been censored out by BN.com's rank reporting strategy. To address this asymmetry, we remove those books with ranks above 650,000 at Amazon.com. More importantly, removing these books serves to remove books for which the ranks are updated very infrequently. As we argued earlier, for books with very high ranks, the ranks no longer represent a snap shot of current sales. Due to this, the sales could have preceded the posting of reviews on the site, in which case we would want to avoid concluding that customer

reviews had any causal relationship to sales. The final sample contains 2394 observations, 1093 of which have reviews posted at both sites.

Table 2 presents the summary statistics for our data. The average sales ranks and the average prices in the sample are very similar across the two sites. Most of the books have a promised delivery of 24 hours (96% at Amazon and 88% at BN.com). The two notable differences across the two sites are: 1) BN.com prices are significantly higher (as can be shown in a paired t-test), 2) Amazon has more reviews than BN.com.

3. The reviewing process and the distribution of reviews

In this section, we provide information about the characteristics of reviews at Amazon.com and Barnesandnoble.com. We then compare the differences in preferences of reviewers across different categories of books across the two sites.

Table 3 presents the cumulative distribution function on the number of reviews across the two sites. As is expected from the summary information, BN.com has a much higher fraction of books with zero reviews compared to Amazon (54.22% versus 12.61%). In fact, the median of the distribution on the number of reviews is 11 on Amazon, and is 0 on BN.com. However, both sites contain a few books with an enormous number of reviews. Interestingly, the two most reviewed books were part of the “Harry Potter” series (by J.K. Rowling): *Harry Potter and the Goblet of Fire* with 4457 reviews on Amazon, and *Harry Potter and the Prisoner of Azkaban* with 956 reviews on BN.com.

It is interesting to compare the frequency with which reviews are posted on Amazon and BN.com to review frequencies in other contexts. For example, Resnick and Zeckhauser (2003) find that over half of buyers on Ebay.com provide some feedback on a completed transaction. In contrast to Ebay, where a transactor can post feedback once and only once per completed transaction, the number of reviews at Amazon.com and Barnesandnoble.com may, in principle, be unrelated to those sites' past sales of the books. Customers who purchased a book elsewhere could post a review, and it is fairly simple for customers to post multiple reviews of a book. Nonetheless, these data suggest that review posting is, relative to feedback provision at Ebay, quite rare. For example, Amazon.com reported that its cumulative pre-orders of *Harry Potter and the Goblet of Fire* totaled 350,000 one minute *before the book's release* in July of 2000. It continued to be on the USA Today national top-ten bestseller list for the next two and half years, so it is likely that Amazon.com sold many more copies of the book after its release. The 4457 reviews of the book posted on the site, then, are very small in comparison to the site's overall sales of the book.

Next, we present the results on the distribution of star ratings in our sample, conditioning on a book having non-zero reviews on both sites. As Table 4 demonstrates, the average star ratings on both sites are quite high. It is interesting to compare the distribution of reviews in this paper to the distributions found in other contexts. For example, Resnick and Zeckhauser (2003) find that 48.3% of buyers on Ebay.com provide no feedback on transactions, 51.2% provide positive feedback, and only 0.5% provide negative or neutral feedback. In this sense, the reviews in this paper have a lot more variance in ratings than the feedback on Ebay.com. There are a number of reasons that can be used to explain this difference, including the fact that on Ebay both sellers and buyers rate each other, which can

result in an incentive to post positive reviews by the buyer that are in turn reciprocated by the seller. Godes and Mayzlin (2003) find that in their sample of online conversations about TV shows, within the sub-sample where conversations could be described as either negative or positive, about 70% of posts were in fact positive. Thus, in all three settings, despite a predominance of positive reviews, there is some variance on the valence of reviews.

In addition, the reviews on BN.com are significantly more positive than the reviews on Amazon.com. An implication of this may be that consumers may be more skeptical when reading a 5-star review on BN.com, compared to a 5-star review on Amazon, which would imply that in our estimation we should account differentially for the effect of star ratings on the two sites. However, despite this general upwards bias, a significant number of reviews have 1 – 3 stars.

Beyond the ratings given by the reviewers, there might be additional information contained in the message text. Unfortunately, reading the reviews is an extremely costly task, and the measures obtained are very noisy as is shown by Godes and Mayzlin (2003). However, one relatively cost-effective measure of the review text is the length (total number of typed characters) contained in the review. A priori, it is not completely clear how to interpret this measure. One possibility is that a longer review represents more effort on the part of the reviewer. Another possibility is that a longer explanation is required to support a “mixed” review. We find partial support for the latter interpretation: Table 5 shows that, at both sites, 1-star and 5-star reviews are much shorter than 2-star, 3-star, and 4-star reviews.

Another pattern that emerges is that Amazon.com reviewers post longer reviews at all star levels than do their peers at Bn.com. This can be due to several reasons, such as Amazon’s ability to elicit more nuanced reviews from its consumers.

It is important for the site comparison methodology in our main analysis that the populations served by Barnesandnoble.com and by Amazon.com do not have very different preferences. That is, suppose that BN.com customers prefer fiction to non-fiction, while Amazon customers have opposite preferences. We would observe higher ratings and higher market shares for fiction books on Amazon, and higher ratings and higher market shares for non-fiction books and BN.com. However, the inferred link between ratings and sales would be essentially due to differences in preferences across the users of the two sites. To rule out the alternative hypothesis that underlying differences in customer preferences across the two sites are driving both differences in reviewer behavior and market shares, we need to demonstrate similar reviewer preferences for categories across the two sites.

In order to do this, we collect data on book subjects. The book's subject is in most cases classified using the system provided by Book Industry Standards and Communications (BISAC) and is available on Bowker's Global Books in Print.com.⁷ In cases where the BISAC subject was not available, we used the subject classification on Amazon. Further, using the original sample of 6429 titles, we aggregated the subjects into broader categories. The complete classification is available in the Appendix. In Table 6, we present the results of standard deviation normalization for each category. The standard deviation normalization for category j at site i is defined as the $(\text{mean star rating for books in category } j \text{ at site } i - \text{overall mean star rating across books at site } i) / \text{overall standard deviation of star ratings for books at site } i$. In constructing this measure, we use only the sample of books that have non-zero reviews at both sites.

⁷ In fact, each book may contain up to 6 subjects. We used first subject only in this study.

Reviewing patterns for different categories of books are remarkably similar across sites. In particular, we find that the signs for standard deviation normalizations are identical for all categories and the magnitudes are similar across the two sites for most categories. On both sites, for example, juvenile fiction is the highest rated category. That is, reviews posted for books in the juvenile fiction category are typically very positive on both sites. On both sites, the least liked books are in the “serious non-fiction” category. This demonstrates a lack of obvious differences in preferences across the users of the two sites.

4. The effect of reviews on sales

In this section, we examine the relationship between a book’s customer reviews and its market share across sites. Our basic specification is fairly simple. We measure how the difference in sales ranks between Amazon.com and Barnesandnoble.com is related to various measures of customer reviews. Equation 1 below gives the basic specification.

$$\ln(\text{rank}_i^A) - \ln(\text{rank}_i^B) = \beta^A \ln(P_i^A) + \beta^B \ln(P_i^B) + X\Gamma + S\Delta + \varepsilon_i \quad (1)$$

where rank denotes the sales rank, the superscripts A and B refer to Amazon.com and Barnesandnoble.com respectively and the subscript i indexes the book title. P denotes price. X denotes the matrix of review variables of interest. S is a matrix of dummy variables summarizing the shipping times promised by each website for each book. For example, the majority of books in our sample are categorized as “usually ship within 24 hours” at both sites. However, Amazon.com and BarnesandNoble.com use other shipping categories such

as “usually ships in 2-3 days” or “Special order: usually ships in 1-2 weeks.” For each book, S has a 1 for the promised ship time category at Amazon.com and a 1 for the promised ship time category for that book at Barnesandnoble.com. Because S is exhaustive of all of the shipping time categories, we do not include a constant term in the regression. In the interest of space, we don’t present the parameters of Δ in the tables.

Table 7 presents results for the full sample of 2394 observations. Column one presents the results for a regression in which no review variables are included, only prices at both sites and the shipping dummies. The price coefficients reflect a combination of own- and cross-price elasticities at both sites. The price coefficient for Amazon.com is positive and statistically significant, suggesting that, when prices rise, sales ranks at Amazon.com become larger, that is, sales fall. The price coefficient is negative for BN.com. This is as expected; recall that the left hand side variable is $\ln(\text{rank})$ at Amazon.com *minus* $\ln(\text{rank})$ at BN.com. Again, when prices rise at BN.com, sales ranks become larger, that is, sales fall at BN.com relative to Amazon.com. The absolute value of the price coefficient is larger at BN.com, suggesting that sales ranks respond more to prices at BN.com than at Amazon.com. This is consistent with the findings in Chevalier and Goolsbee (2003a) that demand is more elastic at BN.com than Amazon.com. An example will give a general sense of the magnitudes of the price elasticities. Consider a book whose other characteristics led to a sales rank of 500 at both Amazon.com with a price of \$10 at both sites. Increasing the price at Amazon to \$12 would be predicted to change the difference in the log ranks at both sites to 0.28, as for example would occur if the rank at Amazon.com moved to 580 and the rank at BN.com moved to 437. What does that mean for sales? The calibrations described above that extend the results from Schnapp and Allwine (2001), suggest that, in the example above, Amazon’s

sales of the book would fall from approximately 145 per week to 129 per week, while BN.com's sales would rise from approximately 21 units per week to 24.

Column 2 includes measures of the total number of reviews for each book. The variables that are included include the natural log of the total number of reviews at Amazon and the natural log of the total number of reviews at BN. These are set equal to zero when the number of reviews equals zero. We also include dummies, one that takes the value one when a title at Amazon.com has no reviews (and zero otherwise) and one that takes the value one when BN.com has no reviews. These results suggest that ranks are lower (sales higher) at Amazon when Amazon has more reviews, and that ranks are lower (sales higher) at BN.com when BN.com has more reviews. This is consistent with evidence from a different data sample in Chevalier and Goolsbee (2003b). The magnitudes are non-trivial. Consider a book with no reviews at either site whose price and other characteristics would suggest a sales rank of 500 at both sites. The posting of an additional 3 reviews at Amazon.com, if it didn't alter the sales rank at BN.com, would be expected to lower the sales rank to number 327, implying incremental sales of approximately 57 books per week.

The specification in Column 2, however, might be somewhat misleading in that obviously, not all reviews are created equal. As the summary data in Section 3 showed, reviews are, on average quite enthusiastic, with at least half of the reviews being 5 stars on both sites. Thus, by including the number of reviews in Column 2, but omitting their content, we are implicitly measuring the effect, on average, of new favorable reviews being posted. Column 3 of Table 7 improves upon this specification, by including the average star value of the book's customer reviews at each site in the regression. Note that the sign of the coefficient

on the Amazon no reviews dummy changes between Column 2 and Column 3, which may at first appear surprising. However, the sign on the coefficient is intuitive once we consider the fact that a book that gets its first review also experiences a change in its average star rating. Suppose that the relative prices and other factors lead the book to have a rank of 500 at both sites, and the book has no reviews on either site. If a book gets one Amazon review with 1, 2 or 3 stars, its rank on Amazon will rise, assuming that its rank on BN.com stays constant. If, on the other hand, it gets a positive review: 4 or 5 stars, its rank on Amazon will fall. As expected, for both sites, the coefficients for the average star value suggest that sales improve when books are rated more highly, but the effect is statistically insignificant for BN.com. To illustrate the magnitude of the effect, consider a book with four 5-star reviews at both Amazon.com and BN.com and a rank of 500 at both sites. Now imagine that one of the 5 star reviews at Amazon.com were changed to a 1-star review. The coefficients imply that, if BN.com's ranking of the book were unchanged by this review change, the rank at Amazon.com would be expected to fall to 603, an estimated change in sales of about 20 books per week.

Column 4 focuses on a different way of measuring review quality. The fraction of reviews that are 1 star reviews and the fraction of reviews that are 5 star reviews are included for each site. As expected, the coefficients suggest that 5 star reviews improve sales and 1 star reviews hurt sales in a statistically significant way at Amazon. The coefficient for 1 star reviews for BN.com is of the expected sign and statistically significant at the 6 percent level. However, the coefficient for 5 star reviews is almost zero but of the "wrong" sign. Nonetheless, it is interesting to note that the 1 star reviews have large coefficients in absolute value, relative to the 5 star reviews, indicating that the relatively rare 1 star reviews carry a lot

of weight with consumers. This result also makes sense when one considers the credibility of 1-star and 5-star reviews. After all, the author or other interested party may “hype” his or her own book by publishing glowing reviews on these websites.⁸ While the author can post a large number of meaningless 5 star reviews cheaply, he or she cannot prevent others from posting 1-star reviews.⁹

One might be concerned that the relationship between the number of reviews and sales is not causal, in that more reviews might be the result of more sales, rather than the cause of more sales. We cannot completely rule out this hypothesis, though there are several factors that speak against it. First, recall that ranks are updated frequently and, in principle, do not “average in” sales over more than the most recent month. However, reviews are typically much older than sales. Fewer than 2% of the reviews in our sample were written during the previous two months. Thus, the reviews predate the measured sales temporally.

Recalculating all results in the paper using only data from reviews more than 2 months old leads to almost numerically identical results. Furthermore, it is important to remember that the causal link between lagged sales on a site and current reviews on that site is not very clear. There is no reason why consumers should be posting reviews necessarily at the same spot where they purchased their books. Consumers buying a book at one site might post at the other, and consumers who bought their books at a brick and mortar store might post at either or both sites.

⁸ For one well-publicized example in economics, see Morin(2003).

⁹ One could argue that posting 1 star reviews of competing books could be a reasonable strategy for an author. We acknowledge that this may be true, although it is not at all clear that two books on the same subject, for example, are substitutes rather than complements.

However, while it might be reasonable to argue that the correlation between the number of reviews and book market shares is spurious, it is harder to argue that the correlation between the star-rating of the reviews and book market shares is spurious. Recall first that the reviews are “stale” relative to the sales ranks; the ranks reflect current sales, the reviews are the opinions of past buyers. Thus, an argument of reverse causality must proceed as follows: even controlling for the number of reviews across sites, differences in the star ratings of the reviews across sites do not cause subsequent movements in sales. Rather, they reflect differences in the preferences of the “barnesandnoble.com user population” and the “Amazon.com user population”. These differences in preferences affect both ranks and ratings.

We find the argument that star-ratings and book market shares are spuriously related due to reverse causality implausible. Our results in Section 3 suggest that there is not a difference in the propensity of Barnesandnoble.com and Amazon.com reviewers to “like” one category of book (such as fiction or self-help). Thus, if the correlation between star ratings and sales ranks reflect differences in the overall preferences of the population to which the two booksellers appeal, these differences are quite subtle.

The robustness of the estimates in Table 7 are further examined in Table 8. In particular, in Table 8 we examine only the subsample of 1093 books that have at least one review on each site. We drop the “no review” variables, but measure the impact of number of reviews and star rankings for this subsample. The results are similar to those presented above. However, the coefficient magnitudes and significance levels for the variables measuring star rankings

are somewhat larger, emphasizing the importance of having higher star rankings for this subsample.

Finally, we examine the relationship between review lengths and sales. To do this, we repeat the specification in Table 9, including the natural log of the average length of all of the reviews for each book at each site. The coefficient is positive and statistically significant at Amazon.com, negative and insignificant at Bn.com. This suggests, controlling for the quality of the book, longer reviews depress the site's relative share. We check the robustness of this result by replacing the average star measures with the fraction of 1 star, 2 star, 3 star, 4 star and 5 star reviews. We find these results quite robust.

There are two possible interpretations of this result. The first, which we view as the less likely, is that encouraging longer, more useful, more nuanced reviews is in fact harmful to sales. More likely, however, is that, within each site, the length of the review is correlated with the enthusiasm of the review in ways that are not captured by the star measures. For example, even within the realm of the statistically dominant 5 star reviews, there could be differing degrees of enthusiasm. That is, some "read like" 4.5 star reviews, while some read more like 5-star reviews. The ones that read like 4.5 star reviews might on average be longer since they are more likely to be mixed – to mention the negative as well as positive aspects of the book. We find some evidence for this in our data. Consider the subsample of 1093 books with at least one review at both sites. Within that group, consider the subsample of 5 star reviews. The average length of these 5 star reviews at Amazon.com is 796 characters for books whose average Amazon.com star rating is 4 or greater, and is 849 characters for books whose average Amazon.com rating is less than 4. Similarly, the average review length at

Bn.com is 491 for 5 star reviews for a book for which the average rating is 4 or greater, and 672 for 5 star reviews for a book for which the average rating is less than 4. Assuming that the books with the lower average ratings have the “less enthusiastic” 5 star reviews, this at least suggests that even within the 5-star category, review length is correlated with the reviewer’s level of enthusiasm for the book.

Regardless of the interpretation of the length results, the results do seem to suggest that customers read and respond to the review content at each site. However, longer reviews do not necessarily stimulate sales.

5. Conclusion

We analyze reviewing practices at Amazon.com and Barnesandnoble.com. We find that customer reviews tend to be very positive at both sites, that they are more detailed at Amazon.com, and that the relative popularity of different types of books are very similar across sites. Our regression estimates suggest that the relative market share of a book across the two sites is related to differences across the sites in the number of reviews for the book and in differences across the sites in the average star ranking of the reviews.

This evidence suggests that customer word-of-mouth has a causal impact on consumer purchasing behavior at two Internet retail sites. We believe that this has not been shown before. That customer content impacts sales is certainly a prerequisite for differences in customer content quality to have any impact on differences in revenues or profitability across retailers. Our evidence however, stops short of showing that the retailer profits from providing such content. For example, there is nothing in our evidence that shows that

customer reviews do not merely move sales around across books within a site. Since Amazon has many more reviewers than rivals, its reviews are on average quite lengthy, and its reviews are on average quite positive, it seems plausible to at least speculate that the total number of books sold at Amazon is higher than it would have been absent the provision of customer review features. Further, and more interestingly, our results show that customers certainly behave *as if* the fit between customer and book is improved by using reviews to screen purchases. One interesting extension to this research would be to examine whether improving a customer's satisfaction with his or her purchases affects subsequent customer loyalty.

6. Tables

Table 1: Initial Sample of Books

Source	# Unique titles
Books in Print	3,617
Publisher's Weekly	2,812
Total	6,429

Table 2: The sample is all books in our database for which the most popular format of the book at Amazon is the same as the most popular format of the book at BN.com.

Summary information

Variable	Mean	Std. Dev.	Min	Max
Amazon sales rank	129467.50	169227.30	7	645406
BN.com sales rank	120872.70	156829.50	6	647611
Amazon price	13.96	14.39	3.25	250
BN.com price	15.50	14.73	3.25	250
Amazon no of reviews	61.27	180.27	0	4457
BN.com no of reviews	12.87	44.60	0	956
Shipping Dummies				
Amazon, up to 24 hours	0.959			
Amazon, 2-7 days	0.024			
Amazon, more than a week	0.005			
Amazon, special order	0.012			
BN.com, 24 hours	0.882			
BN.com, more than 24 hours	0.118			
Number of observations	2394			

Table 3: The sample is all books in our database for which the most popular format of the book at Amazon is the same as the most popular format of the book at BN.com.

CDF of the distribution on the number of reviews

Amazon		BN.com	
x	Prob(no of reviews $\leq x$)	X	Prob(no of reviews $\leq x$)
0	12.61	0	54.22
1	22.18	3	64.04
3	34.21	5	70.97
5	40.02	12	81.12
11	50.79	29	90.23
20	60.69	61	95.03
37	70.47	956	100
64	80.16		
146	90.02		
280	95.03		
4457	100		

Table 4: The sample is as in Tables 2-3 with the additional restriction that non-zero reviews have been posted at both sites.

The distribution of stars

Amazon		Barnes and Noble.com	
Star Rating	Percentage	Star Rating	Percentage
1 star	8.97	1 star	3.44
2 stars	7.53	2 stars	4.07
3 stars	10.56	3 stars	6.00
4 stars	19.89	4 stars	19.27
5 stars	53.05	5 stars	67.22

Average Rating	4.01 stars	Average Rating	4.45 stars
----------------	------------	----------------	------------

Table 5: Average review length by site and number of stars

	Amazon	Bn.com
1 star reviews	765	558
2 star reviews	916	599
3 star reviews	997	566
4 star reviews	949	577
5 star reviews	812	508
Overall	854	529

Table 6: The sample is as in Tables 2-3 with the additional restriction that non-zero reviews have been posted at both sites.

Standard Deviation Normalization, by category

Category	No of books	Amazon average star rating	BN.com average star rating
Adult Fiction	669	-0.261	-0.169
Adult Non-Fiction	101	0.162	0.103
Do-it-yourself	18	0.290	0.328
Entertainment	8	0.301	0.195
Juvenile	144	0.782	0.557
Language & Arts	20	0.406	0.260
Serious Non-fiction	20	-0.617	-0.697
Self-Improvement	60	0.371	0.130
Social Science	48	0.374	0.304
Travel	5	0.370	0.260

Table 7: This table shows regressions in which each data point is a book sold at both Amazon.com and BarnesandNoble.com. The dependent variable is the difference between the sales rank of the book at Amazon and the sales rank of the book at Barnesand Noble.com. The sample is all books in our database for which the most popular format of the book at Amazon is the same as the most popular format of the book at BN.com.

Dependent variable: $\ln(\text{Amazon sales rank}) - \ln(\text{BN.com sales rank})$

	(1)	(2)	(3)	(4)
Amazon $\ln(\text{price})$	1.574 *** (0.160)	1.568 *** (0.156)	1.564 *** (0.155)	1.549 *** (0.156)
BN $\ln(\text{price})$	-1.821 *** (0.148)	-1.863 *** (0.145)	-1.859 *** (0.144)	-1.845 *** (0.145)
Amazon $\ln(\text{no. of reviews})$		-0.191 *** (0.023)	-0.218 *** (0.024)	-0.208 *** (0.024)
BN $\ln(\text{no. of reviews})$		0.118 *** (0.033)	0.133 *** (0.033)	0.132 *** (0.033)
Amazon no reviews dummy		0.234 *** (0.084)	-0.586 *** (0.187)	0.072 *** (0.109)
BN no reviews dummy		-0.253 *** (0.082)	-0.147 (0.100)	-0.337 ** (0.131)
Amazon average star rating			-0.187 *** (0.038)	
BN average star rating			0.025 (0.017)	
Amazon fraction reviews 5 star				-0.260 *** (0.100)
BN fraction reviews 5 star				-0.127 (0.149)
Amazon fraction reviews 1 star				0.508 ** (0.256)
BN fraction reviews 1 star				-0.884 * (0.467)
No. observations	2394	2394	2394	2394
includes shipping dummies?	y	y	y	y
R-squared	0.088	0.131	0.140	0.139

Table 8: This table shows regressions in which each data point is a book sold at both Amazon.com and BarnesandNoble.com. The dependent variable is the difference between the sales rank of the book at Amazon and the sales rank of the book at Barnesand Noble.com. The sample is all books in our database for which the most popular format of the book at Amazon is the same as the most popular format of the book at BN.com and for which the book has at least one customer review at Amazon.com and at least one customer review at BN.com.

Dependent variable: $\ln(\text{Amazon sales rank}) - \ln(\text{BN.com sales rank})$

	(1)	(2)	(3)	(4)
Amazon $\ln(\text{price})$	2.136 *** (0.339)	2.218 *** (0.332)	2.208 *** (0.327)	2.183 *** (0.328)
BN $\ln(\text{price})$	-2.644 *** (0.291)	-2.640 *** (0.285)	-2.644 *** (0.281)	-2.617 *** (0.282)
Amazon $\ln(\text{no. of reviews})$		-0.336 *** (0.049)	-0.381 *** (0.049)	-0.377 *** (0.050)
BN $\ln(\text{no. of reviews})$		0.221 *** (0.052)	0.241 *** (0.051)	0.243 *** (0.052)
Amazon average star rating			-0.444 *** (0.079)	
BN average star rating			0.133 (0.087)	
Amazon fraction reviews 5 star				-0.723 *** (0.236)
BN fraction reviews 5 star				0.083 (0.188)
Amazon fraction reviews 1 star				1.194 ** (0.505)
BN fraction reviews 1 star				-0.986 * (0.566)
No. observations	1093	1093	1093	1093
includes shipping dummies?	y	y	Y	y
R-squared	0.147	0.184	0.211	0.209

Table 9. The Effect of “disagreement” and review length on book market shares. The sample is the set of books available at both sites with reviews at both sites. Dependent variable $\ln(\text{rank})$ at Amazon minus $\ln(\text{rank})$ at Bn.com.

	(1)	(2)
Amazon $\ln(\text{price})$	2.162 *** (0.328)	2.128 *** (0.326)
BN $\ln(\text{price})$	-2.700 *** (0.280)	-2.673 *** (0.281)
Amazon $\ln(\text{no. of reviews})$	-0.419 *** (0.0502)	-0.415 *** (0.0503)
BN $\ln(\text{no. of reviews})$	0.269 *** (0.0518)	0.269 *** (0.052)
Amazon average star rating	-0.422 *** (0.0791)	
BN average star rating	0.158 * (0.0874)	
Amazon fraction reviews 5 star		-0.464 * (0.242)
BN fraction reviews 5 star		0.110 (0.188)
Amazon fraction reviews 1 star		1.594 ** (0.512)
BN fraction reviews 1 star		-1.067 * (0.562)
Amazon $\ln(\text{average rev length})$	0.555 *** (0.146)	0.580 *** 0.151
Bn $\ln(\text{average rev length})$	-0.0351 (0.0916)	-0.0379 (0.0920)
No. observations	1093	1093
Includes shipping dummies?	Y	Y
R-squared	0.222	0.221

7. Appendix

Criteria to select a random sample of books in print

- 1) Only those entries with “A” for the first letter of the author’s last name.
- 2) Only books published in 1998, 1999, 2000, 2001, 2002.
- 3) Search for each of the following “keyword in title” choices: “the,” “of,” and, “a.”

For keyword in title = “the”; “of”; “and”; “a”;

For publication year=1998-2002;

- a. search hardcover fiction
- b. search softcover fiction
- c. search hardcover nonfiction
- d. search softcover nonfiction

Categorization

BISAC	CATEGORY
FICTION	ADULT FICTION
POETRY	ADULT FICTION
TRUE CRIME	ADULT NON-FICTION
CURRENT EVENTS	ADULT NON-FICTION
BIOGRAPHY & AUTOBIOGRAPHY	ADULT NON-FICTION
RELIGION	ADULT NON-FICTION
HOUSE & HOME	DO IT YOURSELF
CRAFTS & HOBBIES	DO IT YOURSELF
GARDENING	DO IT YOURSELF
FOREIGN LANGUAGE STUDY	DO IT YOURSELF
COOKING	DO IT YOURSELF
GAMES	ENTERTAINMENT
HUMOR	ENTERTAINMENT
PETS	ENTERTAINMENT
JUVENILE NONFICTION	JUVENILE
JUVENILE FICTION	JUVENILE
PHOTOGRAPHY	LANGUAGE & ARTS
ANTIQUES & COLLECTIBLES	LANGUAGE & ARTS
ARCHITECTURE	LANGUAGE & ARTS
MUSIC	LANGUAGE & ARTS
PERFORMING ARTS	LANGUAGE & ARTS
ART	LANGUAGE & ARTS
LANGUAGE ARTS & DISCIPLINES	LANGUAGE & ARTS
LITERARY COLLECTIONS	LANGUAGE & ARTS
LITERARY CRITICISM	LANGUAGE & ARTS
DRAMA	LANGUAGE & ARTS
BUSINESS & ECONOMICS	NON-FICTION SERIOUS
LAW	NON-FICTION SERIOUS
REFERENCE	NON-FICTION SERIOUS
MATHEMATICS	NON-FICTION SERIOUS
SCIENCE	NON-FICTION SERIOUS

TECHNOLOGY	NON-FICTION SERIOUS
COMPUTERS	NON-FICTION SERIOUS
STUDY AIDS	SELF-IMPROVEMENT
EDUCATION	SELF-IMPROVEMENT
SELF-HELP	SELF-IMPROVEMENT
FAMILY & RELATIONSHIPS	SELF-IMPROVEMENT
MEDICAL	SELF-IMPROVEMENT
BODY, MIND & SPIRIT	SELF-IMPROVEMENT
SPORTS & RECREATION	SELF-IMPROVEMENT
HEALTH & FITNESS	SELF-IMPROVEMENT
HISTORY	SOCIAL SCIENCE
PHILOSOPHY	SOCIAL SCIENCE
PSYCHOLOGY	SOCIAL SCIENCE
POLITICAL SCIENCE	SOCIAL SCIENCE
SOCIAL SCIENCE	SOCIAL SCIENCE
TRANSPORTATION	TRAVEL
TRAVEL	TRAVEL
NATURE	TRAVEL

References

- Avery, C.; Resnick, P.; Zeckhauser, R. (1999) "The Market for Evaluations," *American Economic Review*. 89 (June), 564-84.
- Chevalier, J. and Goolsbee, A. (2003a) "Measuring prices and price competition online: Amazon.com and BarnesandNoble.com", *Quantitative Marketing and Economics* I(2), forthcoming.
- Chevalier, J. and Goolsbee, A. (2003) "Valuing Internet Retailers: Amazon.com and BarnesandNoble.com", Yale School of Management working paper.
- Coleman, J. S., E. Katz, and H. Menzel (1966). "Medical Innovation: A Diffusion Study," Indianapolis, Indiana: Bobbs-Merrill.
- Eliashberg, J. and Shugan, S. "Film critics: Influencers or predictors?" *Journal of Marketing*, Vol. 61, No. 2 (April 1997), 68-78.
- Ellison, G.; Fudenberg, D. (1995) "Word-of-Mouth Communication and Social Learning," *The Quarterly Journal of Economics*. 110 (1), 93-125.
- Fingar, P.; Kumar, H.; Sharma, T. (2000) *Enterprise E-Commerce*, Meghan-Kiffer Press.
- Foster, A. and M. Rosenzweig (1995, December). "Learning by doing and learning from others: Human capital and technical change in agriculture," *Journal of Political Economy*. 103(6), 1176-1210.
- Godes, D. and Mayzlin, D. (2003) "Using Online Conversations to Study Word of Mouth Communication," Yale SOM working paper.
- Mayzlin, D. (2003) "Promotional Chat on the Internet," Yale SOM working paper.
- McWilliam, G. (2000) "Building Strong Brands through Online Communities," *MIT Sloan Management Review*, 41 (3), 43-54.
- Morin, R. (2003) "Scholar Invents Fans to Answer his Critics," *Washington Post*, 2/1/2003, page C01.
- Resnick, P. and Zeckhauser, R. (2002). "Trust Among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System." *The Economics of the Internet and E-Commerce*. Michael R. Baye, editor. Volume 11 of Advances in Applied Microeconomics. Amsterdam, Elsevier Science.

Schnapp, M. and Allwine, T. (2001) “Mining of book data from Amazon.com”, Presentation at the UCB/SIMS web mining conference,
<http://www.sims.berkeley.edu/resources/affiliates/workshops/webmining/slides/ORA.ppt>

Van den Bulte, C. and G. Lilien (2001). Medical Innovation revisited: Social contagion versus marketing effort. *American Journal of Sociology* 106 (5), 1409–35.