ERIC T. ANDERSON and DUNCAN I. SIMESTER*

The authors document that approximately 5% of product reviews on a large private label retailer's website are submitted by customers with no record of ever purchasing the product they are reviewing. These reviews are significantly more negative than other reviews. They are also less likely to contain expressions describing the fit or feel of the items and more likely to contain linguistic cues associated with deception. More than 12,000 of the firm's best customers have written reviews without confirmed transactions. On average, these customers have each made more than 150 purchases from the firm. This makes it unlikely that the reviews were written by the employees or agents of a competitor and suggests that deceptive reviews may not be limited to the strategic actions of firms. Instead, the phenomenon may be far more prevalent, extending to individual customers who have no financial incentive to influence product ratings.

Keywords: ratings, reviews, deception

Online Supplement: http://dx.doi.org/10.1509/jmr.13.0209

Reviews Without a Purchase: Low Ratings, Loyal Customers, and Deception

In recent years, many Internet retailers have added to the information available to customers by providing mechanisms for customers to post product reviews. In some cases, these reviews have become the primary purpose of the website itself (e.g., Yelp.com, TripAdvisor.com). The increase in product reviews has been matched by an increase in academic interest in word of mouth and the review process (Chevalier and Mayzlin 2006; Godes and Mayzlin 2004, 2009; Lee and Bradlow 2011). Much of this research has focused on why customers write reviews and whether these reviews influence other customers. However, more recently, some of the focus has turned to the study of fraudulent or deceptive reviews (Luca and Zervais 2013; Mayzlin, Dover, and Chevalier 2013).

*Eric T. Anderson is Hartmarx Professor of Marketing, Kellogg School of Management, Northwestern University (e-mail: eric-anderson@kellogg.northwestern.edu). Duncan I. Simester is Nanyang Technological University Professor of Management Science, MIT Sloan School of Management, Massachusetts Institute of Technology (e-mail: simester@mit.edu). The authors thank the apparel retailer that provided the data for this study, as well as seminar participants at University of California, Davis; Stanford University; London Business School; University of Santa Clara; University of Toronto; University of Michigan; Washington University in St. Louis; and the 2013 Yale Customer Insights Conference for many helpful comments. Jacob Goldenberg served as associate editor for this article.

We study product reviews at a prominent private label apparel company. The company's products are only available through the firm's own retail channels; the firm does not allow other retailers to sell its products. The unique features of the data reveal that approximately 5% of the product reviews are written by customers for whom we can find no record of ever purchasing the item. These reviews are significantly more negative on average than the other 95% of reviews, for which there are records that the customer previously purchased the item. They are also significantly less likely to include descriptions of the fit or feel of the garments, which typically can only be evaluated through physical inspection. This is consistent with the interpretation that these reviewers have not purchased the item that they are reviewing. More than 12,000 customers, including some of the firm's highest-volume customers, have written these reviews.

The data enable us to rule out many alternative explanations for why reviews without a confirmed purchase have low ratings, including item differences, reviewer differences, gift recipients, purchases by other customers in the household, customers misidentifying items, changes in item numbers, purchases on secondary markets, unobserved transactions (in retail stores), complaints about non-product-related issues (i.e., shipping or service complaints), and differences

ISSN: 0022-2437 (print), 1547-7193 (electronic)

Journal of Marketing Research, Ahead of Print
DOI: 10.1509/jmr.13.0209

in the timing of the reviews. We caution that even after ruling out this long list of alternative explanations, we cannot conclusively establish that customers never purchased the item (only that we can find no record of a purchase). However, any alternative explanation would need to explain not only why we do not observe a purchase but also why these reviews have low ratings and why there are significant differences in the content of the review text.

We are also able to replicate the low rating effect using a sample of reviews from Amazon.com ("Amazon" for brevity hereinafter). Amazon allows reviewers to add an "Amazon Verified Purchase" tag to their reviews if it can verify that the reviewer purchased the item through Amazon. As a result, reviews without this tag are less likely to have a corresponding purchase than reviews with this tag (although at least some of the reviews without the tag are for items purchased from other retailers). The reviews without the Amazon Verified Purchase tag exhibit the same low rating effect as the reviews from the apparel retailer that we study. We conclude that the low rating effect seems to be a robust effect that generalizes beyond the retailer and the apparel category that we examine in this article.

Product reviews at this retailer are submitted through the company's website. Reviews can only be submitted by registered users, and the information provided in the registration process enables the firm to link the identity of the reviewer to the customer's unique account key, which is the same account key used in the company's transaction data. Registered customers can post a review for any item and are not restricted to posting reviews only for items they have purchased. All the reviewers in our sample are registered users on the website and have purchased from the company through its retail stores, website, or catalogs. A third party screens the reviews for inappropriate content, such as vulgar language or mentions of a competitor. There are no other screening mechanisms on the reviews.

We provide two direct measures indicating that at least some of the reviews without confirmed transactions may be deceptive. First, we identify a sample of reviews in which the reviewers explicitly claim in their review comments that they have purchased the item from the firm. Yet the evidence suggests that at least some of these customers never purchased the item in question. Second, recent research in the psycholinguistics literature stream has identified linguistic cues that indicate when a message is more likely to be deceptive, and we find that the textual comments in the reviews without confirmed transactions exhibit many of these characteristics.

In Figure 1, we provide an example of a review that exhibits linguistic characteristics associated with deception. Perhaps the strongest cue associated with deception is the number of words: deceptive messages tend to be longer. They are also more likely to contain details unrelated to the product (e.g., "I also remember when everything was made in America"), and these details often mention the reviewer's family (e.g., "My dad used to take me when we were young to the original store down the hill"). Other indicators of deception include the use of shorter words and multiple exclamation points.

Previous research on deception in product reviews has largely investigated retailers selling third-party branded products (e.g., Amazon) or independent websites that provide information about third-party branded products (e.g., Zagat.com, TripAdvisor.com). What makes the findings in this study particularly surprising is that the product reviews in this setting are for a single apparel retailer's own private label products. As a result, the strategic incentives to distort reviews are different. A hotel benefits from (deceptively) posting positive reviews about its own property and negative reviews about competing properties on TripAdvisor.com to encourage substitution to its own property (see, e.g., Dellarocas 2006; Luca and Zervais 2013; Mayzlin 2006; Mayzlin, Dover, and Chevalier 2013). However, in the apparel market, the proliferation of items and competitors means that compared with the hotel industry, there are much weaker incentives to write a negative review about a single competitor's product. The firm that we study has hundreds of competitors, and each of the firms sell thousands of products. Because sales are so dispersed, a negative review on a product may lower sales at this firm but have negligible impact on a competitor.

Figure 1

EXAMPLE OF A REVIEW EXHIBITING LINGUISTIC CHARACTERISTICS ASSOCIATED WITH DECEPTION

Over 80 words	I have been shopping at here since I was very young. *My dad used to take me when we were young to the original store down the hill. I also remember when everything was made in America.* I recently bought gloves for my wife that she loves. More recently I bought the same gloves for myself and I can honestly say, "I am totally disappointed"! I will be returning the gloves. My gloves ARE NOT WATER PROOF !!!! They are not the same the same gloves !!! Too bad.	Details unrelated to the product, often referring to the reviewer's family
	Multiple exclamation points	

Notes: This example is based on an actual review. Unimportant details have been modified to protect the identity of the retailer.

Another distinctive feature of the data is that the distortion in the ratings is asymmetric. Although we observe an increase in the frequency of low ratings among reviews without confirmed transactions, there is no evidence of an increase in high ratings. This feature contrasts with previous work, which has found evidence that deceptive reviews on travel sites increase the thickness of both tails in the rating distribution (Luca and Zervais 2013; Mayzlin, Dover, and Chevalier 2013).

The primary contribution of this article is to present evidence that some reviewers write reviews without purchasing the products. We document that the ratings are systematically lower and text comments are significantly different for these reviews. In addition, we show that these reviewers are some of the firm's best customers. The article and accompanying Web Appendix present a wide range of robustness checks for these results. The data are not well-suited to pinpointing why customers might write a review for a product they have not purchased or why those reviews are more likely to be negative. We propose three possible explanations and present initial evidence to investigate these explanations. The explanation that is most consistent with the data is that these loyal customers are acting as self-appointed brand managers. When browsing through the company's website, they observe some products that they do not expect the firm to sell (often new or niche products), and this provokes them to give feedback to the firm. The review process provides a convenient mechanism for them to provide this feedback. We also investigate the possibility that these reviewers are upset customers (although the data do not support this explanation) or that they are trying to enhance their social status. We hope that the findings stimulate other researchers to further investigate these explanations using additional sources of data.

Very few customers write reviews—in the current case, approximately 1.5% of the firm's customers. Reviews without confirmed transactions are written by only 6% of all reviewers. In other words, for every 1,000 customers of this firm, only approximately 15 have ever written a review of the firm's products, and of these, only 1 has written a review without a confirmed transaction (i.e., only 1 in 1,000 customers). We should perhaps not be surprised to observe 1 out of a sample of 1,000 engaging in unexpected behavior. What is concerning is that the reviews written by these 15 customers can influence the behavior of the other 985 customers. This is evident in the data; we show that lower ratings in a review are associated with reduced demand for that product over the next 12 months.

The article proceeds as follows. In the next section, we review the related literature. Then, we describe the data and compare the product ratings and text comments of reviews with and without confirmed purchases. Next, we present evidence indicating that reviews without confirmed transactions contain cues consistent with deception. We rule out several alternative explanations for the low rating effect and also replicate the effect using a sample of book reviews from Amazon. Then, we describe who writes reviews without confirmed transactions and investigate several explanations for why a customer would write a review without having purchased the product. We also present evidence that the low rating effect causes customers not to purchase products that they would otherwise purchase.

LITERATURE REVIEW

The article contributes to the increasing stream of theoretical and empirical work on deceptive reviews. Two articles highlight the theoretical research to date: Mayzlin (2006) and Dellarocas (2006). Mayzlin (2006) studies the incentives of firms to exploit the anonymity of online communities by supplying chat or reviews that promote their products. Her model yields a unique equilibrium whereby promotional chat remains credible (and informative) despite the distortions from deceptive messages. A key element of this model is that inserting deceptive messages is costly to the firm, which means that it is not optimal to produce high volumes of these messages. Although the system continues to be informative, the information content is diminished by the noise introduced by the deception. As result, there is a welfare loss because consumers make less optimal choices. This threat of welfare loss has led to occasional intervention by regulators.[1]

Dellarocas (2006) reports a slightly different result. He describes conditions in which the number of deceptive messages increases with the quality of the firms. This can yield outcomes in which there is better separation between high- and low-quality firms, potentially leading to more informed customer decisions. Social welfare may still be reduced by the presence of deceptive messages if it is costly for the firms to produce them. However, the firms, which must keep up with their competitors, bear the cost of the deception instead of the customers.

The empirical work on deceptive reviews can be traced back to the extensive psychological research on deception (for meta-analyses summarizing this research, see DePaulo et al. 2003; Zuckerman and Driver 1985). The psychological research has often focused on identifying verbal and nonverbal cues that can be used to detect deception in face-to-face communications. However, in electronic and computer-mediated settings, the audience typically does not have access to the same rich array of cues to use to detect deceptions. For example, research has shown that humans are, in general, less accurate at detecting deception using visible cues than using audible cues (Bond and DePaulo 2006). As a result, researchers have observed that deception detection in electronic media is often far more difficult than in face-to-face settings (see, e.g., Donath 1999), which has led to a fast-growing literature stream studying deception detection in electronic media. This includes research in the computer science and machine learning fields that is developing and validating automated deception classifiers for use in the identification of fake reviews (for recent examples, see Jindal and Liu 2007; Mukherjee, Liu, and Glance 2012; Ott et al. 2011).

More closely related to this article is research on the linguistic characteristics of deceptive messages. This includes several studies comparing the linguistic characteristics of text submitted by study participants who are instructed to write either accurate or deceptive text (see, e.g., Zhou 2005; Zhou et al. 2004). Other studies have compared the text of financial disclosures from companies whose filings were later discovered to be fraudulent with filings in which there was no subsequent evidence of fraud (Humphreys et al. 2011). In addition, two studies compare deceptive travel reviews with actual travel reviews. Yoo and Gretzel (2009) obtained 42 deceptive reviews of a Marriott hotel from students in a tourism marketing class and compared them with 40 actual reviews for the hotel posted on TripAdvisor.com. Similarly, Ott et al. (2011) obtained 20 deceptive opinions for each of 20 Chicago-area hotels using Amazon's Mechanical Turk and compared them with 20 TripAdvisor.com reviews for the same hotels. Other studies have compared the content of e-mails (Zhou, Burgoon, and Twitchell 2003), instant messages (Zhou 2005), and online dating profiles (Toma and Hancock 2012). Collectively, these studies yield

[1]Mayzlin, Dover, and Chevalier (2013) cite examples of intervention by both the U.S. Federal Trade Commission and the U.K. Advertising Standards Authority. In September 2013, the New York State Attorney General reached a $350 million settlement with 19 companies that agreed to stop writing fake reviews (Clark 2013).

a series of linguistic cues indicating when a review may be deceptive. We employ these cues in our subsequent analysis.

Several studies have attempted to detect deception in online product reviews without the aid of a constructed sample of deceptive reviews. Wu et al. (2010) evaluate hotel reviews in Ireland by comparing whether positive reviews from reviewers who have posted no other reviews (which they label "positive singletons") distort hotel rankings. Luca and Zervais (2013) use the fraud filter on Yelp.com to distinguish reviews that are likely to be fraudulent. Other authors have used distortions in the patterns of customer feedback on the helpfulness of reviews (see, e.g., Hsu, Khabiri, and Caverlee 2009; Kornish 2009; O'Mahony and Smyth 2009).

A particularly clever recent study compared ratings of 3,082 U.S. hotels on TripAdvisor.com and Expedia.com (Mayzlin, Dover, and Chevalier 2013). Unlike TripAdvisor.com, Expedia.com is a website that reserves hotel stays and thus is able to require a customer to have actually reserved at least one night in a hotel within the prior six months before the customer can post a review. This also links the review to a transaction, making the reviewer's identity more verifiable to the website. In contrast, TripAdvisor.com does not impose the same requirements, which greatly lowers the cost of submitting fake reviews. The key findings are that the distribution of reviews on TripAdvisor.com contains more weight in both extreme tails.

In both prior theoretical research (Dellarocas 2006; Mayzlin 2006) and prior empirical research, the primary focus is on strategic manipulation of reviews by competing firms. For example, Mayzlin, Dover, and Chevalier (2013) show that positive inflation in reviews is greater for hotels that have a greater incentive to inflate their ratings. Similarly, negative ratings are more pronounced at hotels that compete with those hotels. An important distinction we show in this article is that the low ratings in reviews without confirmed transactions are unlikely to be attributable to strategic actions by a competing retailer. Instead, we observe the strongest effects among individual reviewers who purchase a large number of products. This finding has the important implication of broadening the scope of the manipulation of reviews beyond firms that have clear strategic motivations to include individual customers whose motivations seem to be solely intrinsic.

One reason there has been so much recent interest in deceptive reviews is that there is now strong evidence that the reviews matter. For example, Chevalier and Mayzlin (2006) examine how online book reviews at Amazon and BarnesandNoble.com affect book sales. There is evidence not only that positive recommendations and higher ratings lead to higher sales but also that the effect is asymmetric: the negative impact of low ratings is greater than the positive impact of high ratings, which amplifies the importance of any distortion that leads to more negative ratings. This includes our finding that reviews without confirmed transactions are more likely to have low product ratings, without any offsetting increase in the frequency of high ratings.

In the next section, we provide a description of the data used in the study. We present initial evidence of the low rating effect and show that the text comments in these reviews are less likely to contain words describing the fit or feel of the products.

DATA AND INITIAL FINDINGS

The company that provided the data for this study is a prominent retailer that primarily sells apparel. The products are moderately priced (approximately $40 on average), and past customers return to purchase relatively frequently (1.2 orders containing, on average, 2.4 items per year). Although many competitors sell similar products, the company's products are essentially all private label products that are not sold by competing retailers. Our analysis is greatly simplified by the fact that the firm does not allow other retailers to sell its products. Instead, the products are exclusively sold through this firm's retail channels, which include catalog and Internet channels, together with a small number of retail stores.

The firm invests considerable effort to match customers in its retail stores with customers from its catalog and Internet channels by asking for identifying information at the point of sale and matching customers' credit card numbers. Some of this matching is done for the company by specialized firms that use sophisticated matching algorithms. The company has many years of experience matching household accounts. We subsequently investigate whether imperfections in this process may have contributed to the low rating effect.

The company not only matches customer data but also uses credit card numbers and shipping information to identify which customers share a common household. For example, a husband and wife may both order from the firm. They will each have separate customer numbers but a common household number. When matching the transaction and review information, we do so at the household level to identify whether anyone in the household has purchased the item (and not just whether that customer has purchased the item).

On the firm's website, the only way to submit a review is to click the button on each item's product page inviting reviews for that item. The reviewers provide a product rating on a five-point scale, with 1 being the lowest and 5 the highest rating. Almost all the reviews also include text comments submitted by the reviewers. The retailer also has both telephone and online channels that accept feedback about customer service issues, including shipping or sales tax policies. Despite the availability of these alternative channels, it is possible that customers use the product review mechanism to provide feedback about general customer service issues. We investigate this possibility when evaluating alternative explanations for the findings.

The household transaction data we use in this study are a complete record for all customers who purchased an item within the past five years. We only consider reviews written by customers who have made a purchase in this period. This excludes "phantom" reviewers who have never purchased from the firm as well as some actual customers who have not purchased within that five-year window. From an initial total sample of 330,975 reviews, we are left with a final sample of 325,869 reviews used in the study. For 15,759 of the 325,869 reviews (4.8%), we have no record of the customer purchasing the item (although we do have records of that customer purchasing other items).

In Table 1, we report the average product rating for the reviews with and without a confirmed transaction. The dis-

Table 1
DISTRIBUTION OF PRODUCT RATINGS

	Without a Confirmed Transaction	*With a Confirmed Transaction*	*Difference*	
Average rating	4.07	4.33	–.26*	(.01)
Rating = 1	10.66%	5.28%	5.38%*	(.19%)
Rating = 2	6.99%	5.40%	1.59%*	(.19%)
Rating = 3	8.01%	6.47%	1.53%*	(.20%)
Rating = 4	13.83%	16.96%	–3.13%*	(.31%)
Rating = 5	60.51%	65.89%	–5.38%*	(.39%)
Chi-square test			1,156.14*	
Kullback–Leibler divergence			.0259	

*$p < .01$.

Notes: The table reports the average product ratings for reviews with and without a confirmed transaction. The sample sizes are 15,759 (reviews without a confirmed transaction) and 310,110 (reviews with a confirmed transaction). Standard errors are in parentheses.

tribution of reviews without confirmed transactions includes a significantly higher proportion of negative reviews. In particular, there are twice as many reviews with the lowest rating (10.66%) among the reviews without confirmed transactions as for reviews with confirmed transactions (5.28%). We report the Kullback–Leibler divergence with a chi-square test of whether the distributions of product ratings (for items with and without confirmed transactions) are equivalent. The chi-square test statistic confirms that the difference between the distributions is highly significant.

In the Web Appendix, we replicate these findings using a multivariate approach. Specifically, we estimate models in which the dependent variable either measures whether a review has a rating equal to one (a logistic regression model) or measures the product rating itself (ordinary least squares [OLS]). We include variables to explicitly control for the reviewer's characteristics, the item's characteristics, the date of the review, and other characteristics of the review.[2] In addition, we report fixed-effects models, using fixed effects for the item, the reviewer, or the date of the review. The finding that reviews without confirmed transactions have systematically lower ratings remains robust under all these replications.

We argue that many of the reviews for which we cannot find a confirmed transaction were written by reviewers who never purchased the item. However, to support this interpretation, we need to rule out a wide range of alternative explanations. We present this analysis in the "Ruling Out Alternative Explanations" section. Our next set of results helps us identify differences in the text comments that accompany the review. We begin by focusing on whether the text includes a discussion of the fit or feel of the product.

Comments About Fit and Feel

If reviewers never purchased the items they are reviewing, we might expect their reviews to contain fewer references to product features that can only be obtained through physical inspection of the items. For example, reviewers can generally only assess if a material is "soft" or if the fit is "tight" by physically inspecting the item. In Table 2, we compare the frequency with which customers use expressions to describe an item's fit or feel. We obtained these expressions through inspection of a subsample of the actual reviews.[3] To validate the text strings, we used a sample of 500 randomly selected reviews and asked coders, "Does the reviewer comment on the physical fit of the product?" The recall and precision of the "fit" text analysis are 82% and 87%, respectively.[4] We also asked the coders whether the reviewers commented on the "physical feel" of the items. The recall and precision for the "feel" text analysis are 92% and 93%, respectively (for detailed findings, see the Web Appendix).

The findings reveal a pattern: reviews without confirmed transactions are consistently less likely to include these expressions,[5] which is consistent with these reviewers not having physical possession of the items. In the Web Appendix, we repeat this analysis using a series of robustness checks. In particular, we compare the findings when separately examining reviews at each rating level. This controls for the valence of the review. We also repeat the analysis when controlling for the alternative explanations that we identify in the "Ruling Out Alternative Explanations" section.

Summary

We compared the distribution of product ratings for reviews with and without confirmed transactions. The reviews without confirmed transactions have twice as many ratings of 1 (the lowest rating). A comparison of the text comments reveals that the reviews without confirmed transactions are also less likely to contain expressions describing the fit or feel of the items. In the next section, we search for evidence that some of the reviews without confirmed transactions may be deceptive. We do so by again focusing on the text comments in the reviews.

[2]The Web Appendix reports definitions of these variables with summary statistics and pairwise correlations.

[3]The "fit" text strings included "tight," "loose," "small," "big," "long," "narrow," "fit," "fitting," and "blister." The "feel" text strings included "soft," "cozy," "snug," "heavy," "light," "weight," "smooth," "stiff," "warm," "coarse," "felt," "feels," "comfort," "comfy," "flimsy," "they feel," "it feels," "the feel," and "sturdy."

[4]In the pattern recognition literature stream, "precision" is defined as the proportion of retrieved instances (from the text analysis) that are correct (according to the coders), and "recall" is the proportion of correct instances (according to the coders) that are retrieved (by the text analysis).

[5]In the next section, we show that reviews without confirmed transactions tend to have more words in their text comments on average. The relative infrequency of fit and feel expressions occurs despite this higher word count.

Table 2
EXPRESSIONS DESCRIBING FIT AND FEEL

	Without a Confirmed Transaction	*With a Confirmed Transaction*	*Difference*	
Any fit words	43.77%	47.81%	–4.04%*	(.41%)
Any feel words	51.60%	55.15%	–3.56%	(.41%)

*$p < .01$.

Notes: The table reports averages for each measure separately for the samples of reviews with and without confirmed transactions. The sample sizes are 15,759 (reviews without a confirmed transaction) and 310,110 (reviews with a confirmed transaction). Standard errors are in parentheses.

IS THERE EVIDENCE OF DECEPTION?

Detecting deception is inherently difficult because the deceiver tries to avoid detection. In the absence of a constructed sample of deceptive observations (reviews), the standard approach to detect deception is the same approach that we use in this article: compare the characteristics of suspicious observations with a sample of observations that are not considered suspicious. We begin by comparing whether the reviews contain linguistic cues commonly used to identify deception. We then repeat the analysis when restricting attention to reviews in which the reviewers stated that they had actually purchased the item.

Reviewers Who Self-Identified That They Purchased the Item

As we discussed previously, there is an extensive literature stream investigating the differences between deceptive and truthful messages. This research has distinguished face-to-face communications from deception in electronic settings, in which receivers do not have access to the same set of verbal and nonverbal cues with which to detect deception. In electronic settings, the focus of deception detection has largely shifted to the linguistic characteristics of the message. Among the most reliable indicators of deception in electronic settings is the number of words used. Researchers have found evidence that deceptive writing contains more words in many settings, including importance rankings (Zhou, Burgoon, and Twitchell 2003), computer-based dyadic messages (Hancock et al. 2005), mock theft experiments (Burgoon et al. 2003), e-mail messages (Zhou, Burgoon, and Twitchell 2003), and 10-K financial statements (Humphreys et al. 2011). In general, explanations for this effect focus on the deceiver's perceived need for more elaborate explanations to make deceptive messages more persuasive.

Another commonly used cue is the length of the words used. Deception is typically considered a more cognitively complex process than merely stating the truth (Newman et al. 2003; Zhou 2005), leading deceivers to use less complex language. The complexity of the language is often measured by the length of the words used, and several studies have reported that deceptive messages are more likely to contain shorter words (Burgoon et al. 2003).

Because it is often difficult for deceivers to create concrete details in their messages, they have a tendency to include details that are unrelated to the focus of the message. For example, in a study of deception in hotel reviews, Ott et al. (2011) report that deceptive reviews are more likely to contain references to the reviewer's family rather than details of the hotel being reviewed. Other indicators of deception reported in hotel reviews include using more exclamation points (!) (Ott et al. 2011).

To evaluate differences in the reviews' text comments, we constructed the following measures and used them to compare reviews with and without confirmed transactions:

- Word count—The number of words in the review.
- Word length—The average number of letters in each word.
- Family—Does the review contain words describing members of the family?
- Repeated exclamation points—Does the review contain repeated exclamation points (!! or !!!)?

We then compared the averages for these measures in the samples of reviews with and without confirmed transactions. Table 3 reports the findings.

The results again indicate significant differences in the content of the text comments. Recall that word count is one of the most commonly used linguistic cues used to detect deception. The word count for the reviews without confirmed transactions is approximately 40% higher than in the reviews with confirmed transactions. We also observe significant ($p < .01$) differences for each of the other linguistic cues.

A possible explanation for the findings is that the reviews without transactions have lower ratings, and the deception cues might be more common on items with lower ratings. The argument that lower ratings may contribute to the distortion cue results seems particularly plausible for the word count and repeated exclamation points results. When reviewers give ratings of 1, they may use more words and/or more exclamation points to express their opinions. To investigate this possibility, we separately repeated the analysis for reviews at each rating level. The Web Appendix presents the findings. We also replicated the findings using a wide range of robustness checks. In all these replications, the word count and repeated exclamation point findings are extremely robust. The family and word length results typically replicate but are somewhat less robust.

Our second measure of deception focuses on whether reviewers claimed they had purchased the item they are reviewing. Merely writing a review without having purchased the item is not necessarily deceptive. However, it *would* be deceptive for reviewers to state that they had purchased an item that they had never purchased. To find reviewers who self-identified that they had purchased the item, we searched in the review comments for text strings indicating that the reviewers were claiming that they had purchased the items.[6] The recall and precision are 83% and 91%, respectively (we report detailed findings in the Web Appendix). The text analysis identified a total of 150,419 reviews in which reviewers stated that they had purchased the item. Of these 150,419 reviews, 7,660 (5.1%) did not have a confirmed transaction. We repeated our comparison

[6]The text strings included "bought," "buy," "purchase," "order," "gave," "I got myself," "I have been looking," "searching," "I waited," "I read," "we got," and "sold." The strings are not case sensitive.

Table 3

INDICATORS THAT A MESSAGE IS DECEPTIVE

	Without a Confirmed Transaction	*With a Confirmed Transaction*	*Difference*	
Word count	70.13	52.00	18.13*	(.33)
Word length	4.110	4.153	–.043*	(.004)
Family	20.74%	18.75%	1.98%*	(.32%)
Repeated exclamation points	6.91%	4.71%	2.20%*	(.18%)

*$p < .01$.

Notes: The table reports averages for each measure separately for the samples of reviews with and without confirmed transactions. The sample sizes are 15,759 (reviews without a confirmed transaction) and 310,110 (reviews with a confirmed transaction). Standard errors are in parentheses.

of both the ratings and the review text using this sample of reviews. Table 4 presents the findings.

When reviewers self-identified that they had purchased the item, we continue to observe both a higher incidence of low ratings among reviews without confirmed transactions and significant differences in the content of the text comments. The reviews without confirmed transactions are less likely to include descriptions of the fit and feel of the garments but tend to contain significantly more words, more mentions of the reviewer's family, and more frequent use of repeated exclamation points.

Summary

We searched for evidence of deception by comparing the text comments in the reviews with and without confirmed transactions. The reviews without confirmed transactions were more likely to contain linguistic cues associated with deception. We also identified a sample of reviews in which reviewers explicitly stated that they had purchased the items. We were able to replicate our earlier findings when restricting attention to reviews in this sample. As we acknowledged at the beginning of this section, it is difficult to find evidence of deception. Therefore, this evidence is best interpreted as indicative but not conclusive. We also emphasize that these differences do not indicate that all of the reviews without confirmed transactions are deceptive.

The restriction to customers who self-identified that they purchased the item also serves another role. By claiming that they had purchased the items, the reviewers explicitly rule out two alternative explanations for why customers might write a review without having purchased the item. First, it is possible that a reviewer could inspect an item without purchasing it. For example, the reviewer may have seen the item worn by a friend or family member. Second, it is also possible that the reviewer may have physically inspected the item in one of the firm's retail stores and then decided not to buy it (which could also explain why the ratings are more negative). Neither of these possibilities is consistent with customers explicitly stating that they had purchased the items. These explanations also do not explain the differences in the content of the text comments.

It is also likely that at least some of the reviewers received the item as a gift, which would explain why we do not observe a transaction for that reviewer. On the one hand, because gift recipients often do not select their gifts, their reviews might also be expected to have lower ratings. On the other hand, it is not clear why gift recipients would be less likely to describe the fit or feel of the products or why they are more likely to include linguistic cues associated with deception. This explanation is also inconsistent with reviewers stating that they had purchased the item. It is possible that some customers who received the item as a gift, perhaps having placed it on a wish list or registry, interpreted this as a "purchase" when they received the item. However, this would be a somewhat unnatural interpretation of a purchase. We conclude that replication of our findings with these customers suggests that the low ratings and differences in the text comments cannot easily be attributed to gift recipients. In the next section, we attempt to rule out a wide range of other explanations for the low rating effect.

Table 4

CUSTOMERS WHO SELF-IDENTIFIED THEY PURCHASED THE ITEM

	Without a Confirmed Transaction	*With a Confirmed Transaction*	*Difference*
Average rating	4.03	4.32	–.29** (.01)
Rating = 1	12.11%	5.84%	6.27%** (.28%)
Rating = 2	7.18%	5.41%	1.77%** (.27%)
Rating = 3	7.39%	6.39%	1.00%** (.29%)
Rating = 4	12.60%	15.72%	–3.12%** (.43%)
Rating = 5	60.72%	66.65%	–5.93%** (.55%)
Chi-square test			660.72**
Kullback–Leibler divergence			.0297
Fit and Feel Analysis			
Any fit words	48.39%	53.97%	–5.57%** (.59%)
Any feel words	52.69%	55.35%	–2.66%** (.58%)
Linguistic Deception Cues			
Word count	83.49	65.71	17.77** (.53)
Word length	4.06	4.08	–.016** (.004)
Family	26.37%	25.14%	1.23%* (.51%)
Repeated exclamation points	7.95%	5.65%	2.30%** (.27%)

*$p < .05$.
**$p < .01$.
Notes: The table reports the average product ratings for reviews with and without a confirmed transaction. The sample includes all of the reviews with the words "bought," "purchased," or "ordered" in the text field. Standard errors are in parentheses. The sample sizes are 7,660 (reviews without a confirmed transaction) and 142,759 (reviews with a confirmed transaction).

RULING OUT ALTERNATIVE EXPLANATIONS

In this section, we investigate several explanations for why we observe lower ratings in reviews without a confirmed transaction. We then establish the robustness of our text analysis. Finally, we replicate the low rating effect using a sample of data from Amazon. Because these robustness checks are so extensive, we summarize the findings in this section and provide a more complete description of the alternative explanations, methodological approach, and results in the Appendix.

The Low Ratings Effect

The first class of alternative explanations includes differences among time periods, products, or reviewers. For example, the items or reviewers in our two samples may be systematically different. If this were true, the low rating effect could be due to a selection problem. We approach this problem using a "within" estimator; we conduct within–time period, within-item, and within-reviewer analyses. The low rating effect survives in all of these separate analyses.

The second class of alternative explanations falls into the category of misclassification. That is, a customer may have purchased the product that he or she reviewed, but we misclassified the review as not having a confirmed purchase. To investigate this possibility, we examine various subsets of the data. For example, we restrict our analysis to customers who live more than 400 miles from the firm's nearest retail store and items for which there are essentially no purchases in the firm's retail stores. This analysis makes it unlikely that the results reflect unobserved purchases through the retail store channel. Similarly, a customer may have obtained the product through a third party, such as eBay or Craigslist. We investigate this possibility by focusing on a product category that is typically not available in secondary markets (underwear). Finally, a customer may accidentally select the wrong product when writing a review. For example, when reviewing a pair of men's pants, the reviewer may have selected the wrong style. To correct for this possibility, we relax our classification rule and link reviews to transactions at the subcategory level. Because subcategories include a wide variety of similar items, this process corrects for this type of customer error. Again, the low ratings result survives each of these robustness checks.

The third class of explanations is that the reviewers may be venting general dissatisfaction with the company through a product review. However, the company offers a variety of ways for customers to provide feedback about service problems, making it less likely that customers will use product reviews to provide feedback about these problems. Nevertheless, we conducted extensive text search to identify complaints related to shipping or customer service. We found no evidence that customers are complaining about these issues through product reviews.

Analysis of the Text Comments

The previous section indicates that reviews without confirmed transactions tend to have significantly fewer words in the text comments describing the fit or feel of the items. We also show that they are more likely to contain linguistic cues associated with deception. As a robustness check, we compared whether these differences in the text survive the different approaches used to control for alternative explanations. We summarize the results in the Web Appendix.

Reviews without confirmed transactions are less likely to contain words associated with the fit or feel of the items, even when controlling for item and reviewer differences. These results also survive matching transactions at the subcategory level, excluding reviewers with store purchases or who live close to a store, and restricting attention to items with few store purchases. The differences are also essentially unchanged when focusing only on the underwear product category, although the reduction in sample size means that the comparison is no longer statistically significant. Finally, the differences survive when we control for the timing of the review (together with other characteristics of the review, the item, and the reviewer) in a logistic regression model.

The replication of the linguistic cues using the same procedures also reveals a robust pattern of results, especially for the word count and repeated exclamation point measures. Recall that the word count measure is one of the most reliable indicators of deception. Reviews without confirmed transactions contain significantly more words under all these replications. The magnitude of the difference in the word count is also essentially unchanged, except when comparing the word count for reviews written by the same reviewer (the within-reviewer analysis). However, this is a relatively conservative test because it eliminates all the between-customer variation.

Other Explanations

We have been able to investigate a wide range of explanations for the differences in reviews with and without confirmed transactions. However, we recognize that ruling out every alternative explanation is not possible. For example, there may be patterns in the data that make our investigations of the alternative explanations incomplete. These patterns include the following:

- Unknown data discrepancies that prevent us from linking a purchase to a review,
- Gift recipients who may describe a gift as a purchase (somewhat unnaturally), and
- Customers who may visit a retail store on vacation even though they do not live close to a store and have never previously purchased in a store.

Although such alternative explanations are possible, we believe that there are several factors that make them (and others) unlikely. First, any unusual patterns in the data must affect a large number of customers. As we discuss subsequently, more than 12,000 customers have written a review without a confirmed transaction. Second, any alternative explanation must not only explain why we do not observe a confirmed transaction but also explain the difference in the product ratings as well as differences in the content of the review text (including both the less frequent use of words describing fit or feel and the increased use of linguistic cues associated with deception). As a final investigation into the robustness of the finding, we next investigate whether the effect replicates using data from Amazon.

Replication of the Low Rating Effect at Amazon

In 2009, Amazon began offering reviewers the option of tagging reviews as an "Amazon Verified Purchase" if the reviewer purchased the item at Amazon.[7] This feature provides an opportunity to replicate our findings using a different retailer and product category.

We selected a sample of 80 books sold by Amazon using an independent random book title generator (www.kitt.net/php/title.php) to generate plausible titles for books. We then searched for these keywords using the advanced search function within Amazon's book department. We restricted attention to books that had between 80 and 100 reviews and only used books published after September 2009 because this is the first month that we can confirm that Amazon was using the Amazon Verified Purchase tag on its reviews.[8] The 80 books include a range of genres, including adult, religion, teen fiction, history, cookbooks, self-help, romance, and humor.

The sample of 80 books had a total of 7,219 reviews, averaging 90.2 reviews per book. This sample included an average of 52.7 reviews tagged as an Amazon Verified Purchase and 37.6 that were not tagged. In Table 5, we report the average rating and the distribution of ratings for these two samples of reviews. We note that the low rating effect is replicated using these reviews from a separate retailer in a different category. The magnitude of the effect is similar to the findings reported in Table 1, with approximately twice

[7]Other firms now use similar tags, including PowerReviews.com, TheUnicycleStore.com, and KaviSkin.com.

[8]There is a reference to the Amazon Verified Purchase tag in a discussion forum on September 20, 2009 (http://www.historicalfictiononline.com/forums/showthread.php?t=2423). We found a second reference to the tag in a different discussion forum on November 29, 2009 (http://www.mobileread.com/forums/showthread.php?t=63708). We also excluded a small number of books for which reviewers had submitted reviews under Amazon's Vine program, in which reviewers are provided with free books in return for submitting reviews.

Table 5
REPLICATION USING BOOK REVIEWS AT AMAZON

	Not an Amazon Verified Purchase	*Amazon Verified Purchase*	*Difference*	
Average rating	4.03	4.25	–.22*	(.03)
Rating = 1	9.38%	4.77%	4.61%*	(.63%)
Rating = 2	6.12%	5.22%	.90%	(.56%)
Rating = 3	10.21%	9.76%	.46%	(.72%)
Rating = 4	21.16%	21.22%	–.06%	(.98%)
Rating = 5	53.13%	59.03%	–5.90%*	(1.18%)
Chi-square test			413.96*	
Kullback–Leibler divergence			.0178	

*$p < .01$.
Notes: The table reports the average product ratings for reviews with and without a confirmed transaction. The sample sizes are 3,006 (reviews without a confirmed transaction) and 4,213 (reviews with a confirmed transaction). Standard errors are in parentheses.

as many ratings equal to 1 among the reviews without a verified Amazon transaction (9.38% vs. 4.77%).[9]

The book market shares several characteristics with the apparel market. Notably, sales are dispersed across a wide range of products and authors. This makes it less likely that the low ratings reflect strategic behavior by competitors. However, we might expect that authors would try to *increase* the average rating of their book(s). If authors inflate the ratings for their books, we would expect an increase in the number of high ratings for reviews without verified transactions. The comparison in Table 5 does not reveal any evidence of this. Rather, we observe the same asymmetry in these data as in Table 1; for the reviews without a verified transaction, there is an increase in the frequency of low ratings and a *decrease* in the frequency of high ratings. A possible explanation for why we do not observe more high ratings among reviews with the Amazon Verified Purchase tag is that the authors (or their confederates) may purchase books from Amazon when submitting favorable reviews to inflate their ratings. An Internet search confirms that some third-party firms advertise that they will submit Amazon Verified Reviews for a fee, which includes the cost of purchasing the book through Amazon.[10]

Another important difference between the apparel results and this replication using the Amazon data is the number of reviews not associated with a confirmed transaction. Recall that in the apparel data, approximately 5% of the reviews are not associated with a confirmed transaction, whereas in the Amazon data, 41.6% of the reviews do not have the Amazon Verified Purchase tag. A simple explanation for this difference is that reviewers can obtain books from many places other than Amazon. In contrast, the apparel sold by the private label retailer can only be purchased through this firm's own retail channels. Because customers can obtain books from other sources, it is likely that at least some of the reviews without the Amazon Verified Purchase tag were written by customers who had purchased the item. However, Amazon's website states that reviewers can only add the verified purchase tag if the firm "can verify the item being reviewed was purchased at Amazon," so it is clear that reviews *without* the tag are less likely to have a corresponding purchase than reviews with this tag. These reviews exhibit the same low rating effect as the reviews from the apparel retailer that we have studied, and the effect is again unlikely to be due to strategic behavior by competitors. We conclude that the low rating effect seems to be a robust effect that generalizes beyond the retailer and the apparel category that we study.

Summary

We have investigated alternative explanations for the lower ratings on reviews without confirmed transactions. The evidence suggests that the low rating effect cannot be attributed to item differences, reviewer differences, gift recipients, purchases by other customers in the household, customers misidentifying items, changes in item numbers, purchases on secondary markets, unobserved transactions (in retail stores), complaints about non-product-related issues (such as shipping or service complaints), or differences in the timing of the reviews. We also use the same procedures to show that these alternative explanations cannot explain the difference in the content of the review text. Finally, using a sample of data from Amazon, we replicate the low rating effect by showing that ratings are lower when reviews do not include the Amazon Verified Purchase tag.

In the next section, we investigate who writes reviews without confirmed transactions. In particular, we evaluate whether the reviews are contributed by the employees or agents of a competitor.

WHO IS IN THE TAIL OF THE TAIL?

We begin this section by investigating how many reviewers wrote reviews without confirmed transactions. We then study which reviewers contributed the low ratings. We conclude by comparing the reviewers' demographic characteristics and historical behavior.

How Many Reviewers Write Reviews Without Confirmed Transactions?

In Table 6, we aggregate the reviews to the reviewer level and group the reviewers according to the number of reviews they have written without confirmed transactions. The findings reveal that more than 94% of reviewers only wrote reviews when they had confirmed transactions. Only 6% of reviewers wrote reviews without confirmed transactions, but this includes more than 12,474 individual reviewers. Of the 15,759 reviews without a confirmed purchase, 12,895 of them (81.8%) were contributed by 11,944 reviewers who wrote just one or two of these reviews.

[9]We also replicated this analysis when controlling for differences between the 80 books. For each book, we calculated the distributions of ratings separately for the reviews tagged versus not tagged as Amazon Verified Purchases. We then compared the difference in these ratings for each book and averaged these differences across the 80 books. This approach is analogous to our control for item differences (see the Appendix). Using this within-book comparison, an average of 5.11% ratings are equal to 1 when the review is tagged as an Amazon Verified Purchases compared with 8.43% when the review does not have this tag. The difference in these averages is statistically significant ($p < .01$).

[10]For example, in April 2013, TheBookplex.com advertised that it charges an administrative fee of $90 for five complete detailed book reviews plus the cost of the books. Reviews with the Amazon Verified Purchase tag can also be purchased from BuyAmazonReviews.com and marketplaces such as Ufiverr.com.

Table 6

HOW MANY REVIEWERS WRITE REVIEWS WITHOUT CONFIRMED TRANSACTIONS?

Number of Reviews Without Confirmed Transactions	*Number of Reviewers*	*Percentage of All Reviewers*	*Percentage of All Reviews Without Confirmed Transactions*
0	200,731	94.15%	—
1	10,993	5.16%	69.76%
2	951	.45%	12.07%
3	249	.12%	4.74%
4	103	.05%	2.61%
5	56	.03%	1.78%
6	28	.01%	1.07%
7	24	.01%	1.07%
8	10	.00%	.51%
9	11	.01%	.63%
10 or more	49	.02%	5.77%

Notes: The table groups reviewers according to the number of reviews they have written without confirmed transactions. The unit of analysis is a reviewer.

Even though most of the reviews without transactions were written by different individual reviewers, it is still possible that the low rating effect is attributable to a small number of reviewers. In Table 7, we report the average rating and proportion of reviews with low ratings when grouping reviewers according to the total number of reviews they have written that have no confirmed transactions. Among the reviews without confirmed transactions, the most negative reviews were written by reviewers who wrote just one of these reviews. We conclude that the low rating effect is attributable to thousands of individual reviewers.

Another finding of interest in Table 7 is that for the 11,944 reviewers (10,993 + 951) who wrote a total of either one or two reviews without confirmed transactions, there is no evidence of low ratings in their reviews when they had purchased the item. When they had a confirmed transaction, these reviewers had the same proportion of low ratings (5.79% and 5.71%) as the 200,731 reviewers who had confirmed transactions for all of their reviews (5.76%). This finding further confirms that the effect cannot be attributed to reviewer differences.

Who Writes Reviews Without Confirmed Transactions?

In Table 8, we summarize the reviewers' purchasing characteristics with a series of demographic variables. We report definitions of these variables, summary statistics, and pairwise correlations in the Web Appendix. We compare reviewers who only wrote reviews with confirmed transactions with reviewers who wrote at least one review without a confirmed transaction. As a benchmark, we also include findings for customers who have never written a review. At the request of the retailer, the age, estimated home value, and estimated household income measures are indexed to 100% for customers who only wrote reviews with confirmed transactions.

We focus first on customers who have written reviews and contrast those who have written at least one review without a confirmed transaction (Table 8, Column 2) with those who have only written reviews with confirmed transactions (Table 8, Column 3). Customers who write reviews without confirmed transactions tend to be younger, have more children in their households, are less likely to be married, and are less likely to have graduate degrees (compared with reviewers who only write reviews with confirmed transactions). They have less expensive homes and lower household incomes. They also tend to be higher-volume purchasers, buying 30% more items even though they have been customers for a slightly shorter period. The average price they pay is identical to the other reviewers, although this price is more likely to be a discounted price. They also write more than twice as many reviews.

In the Web Appendix, we report the findings from a logistic regression model predicting which reviewers wrote at least one review without a confirmed purchase. Several of the reviewer characteristics are accurate predictors, including when they write their reviews, how many reviews they write, how many items they purchase, the price of the items,

Table 7

RATINGS BY NUMBER OF REVIEWS WITHOUT CONFIRMED TRANSACTIONS

Number of Reviews Without Confirmed Transactions	*Average Rating: Without a Confirmed Transaction*		*Average Rating: With a Confirmed Transaction*		*Reviews with Ratings = 1: Without a Confirmed Transaction*		*Reviews with Ratings = 1: With a Confirmed Transaction*		*Sample Size (Number of Reviewers)*
0	—		4.32	(.002)	—		5.76%	(.05%)	200,731
1	3.99	(.01)	4.26	(.02)	12.09%	(.31%)	5.79%	(.31%)	10,993
2	4.11	(.04)	4.28	(.04)	9.62%	(.80%)	5.71%	(.75%)	951
3	4.22	(.06)	4.27	(.06)	6.29%	(1.07%)	4.13%	(.95%)	249
4	4.20	(.09)	4.41	(.07)	8.01%	(2.02%)	2.70%	(.95%)	103
5	4.31	(.11)	4.28	(.12)	6.79%	(2.06%)	5.12%	(1.66%)	56
6	4.42	(.14)	4.56	(.11)	4.76%	(3.19%)	.70%	(.53%)	28
7	4.49	(.13)	4.47	(.13)	3.57%	(2.15%)	1.91%	(1.01%)	24
8	4.20	(.20)	4.47	(.19)	5.00%	(2.76%)	2.78%	(2.78%)	10
9	4.09	(.36)	4.47	(.20)	11.11%	(6.87%)	4.97%	(3.10%)	11
10 or more	4.46	(.09)	4.51	(.08)	4.64%	(1.62%)	3.52%	(1.43%)	49

Notes: The table groups reviewers according to the number of reviews they have written without confirmed transactions. The unit of analysis is a reviewer. Standard errors are in parentheses.

Table 8
DEMOGRAPHICS AND HISTORICAL BEHAVIOR

	Customers Who Have Never Written a Review	*Reviewers with at Least One Review Without a Confirmed Transaction*	*Reviewers Who Have Only Written Reviews with Confirmed Transactions*	*Difference*
Demographics				
Number of children	.50	.59	.49	.11%** (.01)
Married	68.39%	71.95%	73.30%	–1.35%** (.42%)
Age	100.03%	93.47%	100.00%	–6.52%** (.26)
Estimated home value	100.03%	97.94%	100.00%	–2.06%* (.97)
Estimated household income	95.27%	98.64%	100.00%	–1.36%* (.56)
Graduate degree	24.68%	29.70%	31.26%	1.56%** (.44%)
Historical Behavior				
Number of reviews	.00	2.96	1.44	1.53** (.02)
Items purchased	36.10	156.08	119.73	36.35** (1.71)
Average item price	$42.72	$40.99	$40.89	$.11.00 ($.16)
Overall discount received	4.62%	8.76%	7.28%	1.49%** (.08%)
Discount frequency	12.08%	21.59%	17.94%	3.64%** (.17%)
Return rate	12.71%	18.15%	15.63%	2.52%** (.17%)
Years since first order	9.20	11.70	12.52	.82** (.06)

*$p < .05$.
**$p < .01$.
Notes: The table reports averages for each measure separately for the samples of reviews with and without confirmed transactions. Standard errors are in parentheses. The sample sizes for the historical purchasing measures are 12,474 (reviewers: no confirmed transaction) and 200,731 (reviewers: all have a confirmed transaction). The sample sizes for the demographic measures are up to 15% smaller due to missing data for some of these variables. The "Customers Who Have Never Written a Review" sample size is several million (the precise number is confidential). The age, estimated home value, and estimated household income variables are indexed to 100% in the "Reviewers Who Have Only Written Reviews with Confirmed Transactions" sample.

their propensity to purchase on discounts, their return rate, their age, the number of children they have, and whether they are married. However, we caution that the classification table reveals only a very modest improvement in predictive accuracy over a benchmark prediction that none of the reviewers write reviews without a prior transaction.

It is clear that reviewers who write reviews without confirmed purchases are valuable customers. Moreover, the findings seem to confirm that the effect is not due to competitors writing negative reviews to strategically lower quality perceptions for the company's products. If this were the case, we might expect the negative reviews to be concentrated among a handful of reviewers rather than contributed by thousands of individual reviewers. We would also not expect the negative reviewers to have made so many purchases.

Comparing Reviewers with Other Customers

The findings in Table 8 also highlight several differences between reviewers (Columns 2 and 3) and customers who have never written a review (Column 1). If we define a current customer as a customer who has purchased within the past five years, only approximately 1.5% of customers have ever written a review. Reviewers are more likely to be married, have higher household incomes, and have graduate degrees. They also purchase almost four times as many items, have been customers for longer, return more items, and purchase more items at a discount. Although not reported in Table 8, reviewers are also more likely to purchase newly introduced items, items from new categories, and niche items that sell relatively few units. We conclude that the small tail of reviewers is not representative of the other customers who purchase from this firm. In the next section, we investigate explanations as to why a customer might write a review without having purchased the product.

WHY WOULD A CUSTOMER WRITE A REVIEW WITHOUT PURCHASING?

As we have discussed, the primary contribution of this article is to (1) present evidence that some reviewers write reviews without purchasing the products in question, (2) document that the ratings are systematically lower and the text comments are significantly different for these reviews, and (3) verify that these reviews are written by some of the firm's best customers. In this section, we propose three explanations for why a customer would write a review without purchasing. The explanations address both why a customer would write a review and why these reviews tend to have low ratings. We caution that the data are not well-suited to validating these explanations conclusively. Instead, we present initial evidence and hope that the findings stimulate researchers to further investigate these explanations using additional sources of data.

Upset Customers

Our first explanation is that these customers may have experienced a service failure or had some other type of negative interaction with the company. This experience may have prompted the customer to respond by writing a negative review as retribution.[11] We used two approaches to investigate this possibility.

First, we identified text strings that might indicate that the customer is upset or angry with the company.[12] Using our

[11]This explanation is closely related to the psychological phenomenon of negative reciprocity (see, e.g., Eisenberger et al. 2004).

[12]We used the following text strings: "angry," "annoyed," "irritated," "mad," "fuming," "livid," "irate," "furious," "outraged," "infuriated," "upset," "frustrated," "displeased," "aggravated," "exasperated," "maddened," "enraged," "riled," "incensed," "exasperating," "very unhappy," "shame on you," "you owe it to your customer," "order anymore," "driven me," "buying another," and "was the best."

random sample of 500 reviews, we found the recall and precision measures for these text strings to be 80% and 89%, respectively (see the Web Appendix). However, we caution that obtaining reliable measures of recall and precision from a random sample of reviews is difficult because relatively few (.57%) of this firm's reviews seem to have been written by upset customers.

In Figure 2, we report the percentage of reviews that contain at least one of these words for each rating level. For products with a rating of 1, there is almost no difference in the use of these words between reviews with and without confirmed transactions. If anything, customers are more likely to use these words when there is a confirmed transaction. This finding suggests that the customers writing negative reviews without a confirmed transaction are not more upset with the firm than customers writing negative reviews with a confirmed transaction.[13]

Our second approach to investigating this explanation is to compare the change in customers' ordering rates before versus after the review date. If customers are upset with the firm, we would expect a lower rate of subsequent purchases. We control for differences in the rate that customers place orders by calculating each customer's average purchase interval in their previous orders (before the review date). We constructed the following measures:

- Years until next order—Time until the customer places another order (in years).
- Purchase intervals until next order—The number of that customer's average purchase intervals before the customer places another order.
- No subsequent order—Equal to 1 if the customer places no orders after the review date and 0 if otherwise.
- No order in next purchase interval—Equal to 1 if the customer places no orders in the next average purchase interval and 0 if otherwise.
- No order in next year—Equal to 1 if the customer places no orders in the next year and 0 if otherwise.
- More orders in next year versus prior year—Equal to 1 if the customer places more orders in the year after the review date than in the year before and 0 if otherwise.
- More orders in next year versus prior average—Equal to 1 if the customer places more orders in the year after the review date than their average annual purchase rate (before the review date) and 0 if otherwise.

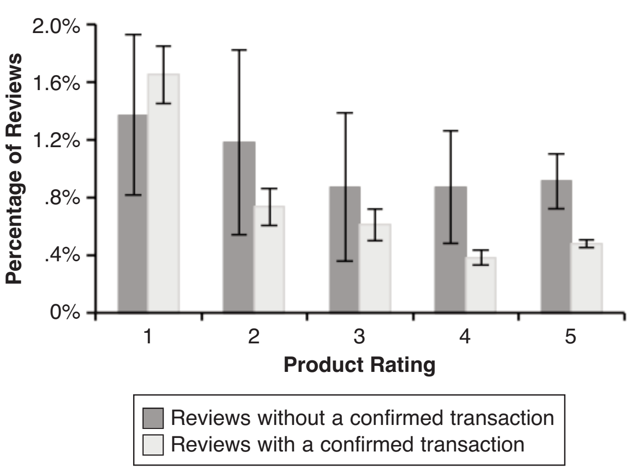
The unit of observation is a reviewer × review date. We report the findings in Table 9, in which we group the observations according to whether the reviewer wrote any reviews on that date without a confirmed transaction. In Table 9, we restrict attention to negative reviews by focusing on observations for which at least one of the reviewer's product ratings on that date was equal to 1. In the Web Appendix, we report the findings that include all the observations.[14] The customers who wrote reviews without a confirmed transaction are more likely to make a subsequent purchase, the interval until their next purchase is shorter, and they are more likely to purchase at a higher rate than in previous periods. This is not what we would expect if the customers were upset with the firm.

It is possible that reviewers may have been upset for some time, so that the preperiod may include some weeks in which reviewers were already upset. Therefore, we replicated the findings (using the more orders in next year vs. prior year measure) when adding an interval between the end of the prior period and the review date. Approximately 75% of customers wrote a review within eight weeks of purchasing the item. Therefore, we repeated the analysis when the preperiod finishes two weeks, four weeks, six weeks, or eight weeks before the review date. The pattern of findings was unchanged. We conclude that the customers who wrote negative reviews without a confirmed purchase seem to be no more upset with the firm than the customers who wrote negative reviews with a confirmed transaction.

Self-Appointed Brand Managers

The second explanation is, in some respects, the reverse of the "upset customers" explanation. It is possible that these customers are acting as "self-appointed brand managers." They are loyal to the brand and want an avenue to provide feedback to the company about how to improve its

Figure 2
ARE THE REVIEWERS UPSET?



Notes: The figure reports the percentage of reviews that included any words associated with upset customers. The sample sizes include 15,759 reviews without and 310,110 reviews with confirmed transactions. The error bars are 95% confidence intervals. For detailed results, see the Web Appendix.

[13]Some might wonder why customers would use these words when they are not upset. We read all the reviews with a rating of 5 (the highest rating) that used these words. This revealed that reviewers sometimes use the words when they are not upset with the firm (e.g., "My boys love these pants and get upset if I have to wash them," "I've been frustrated with pants from other retailers"). Note also that the text strings appear more frequently in *positive* reviews written without (vs. with) a confirmed transaction. This is perhaps consistent with our evidence that these reviews are more likely to contain multiple exclamation points. Reviewers seem to use more expressive words when writing without a confirmed transaction. This discussion highlights the difficulty of obtaining reliable measures of recall and precision for these text strings.

[14]We also include a series of fixed-effects models using each of the seven outcome measures as dependent variables. We include reviewer fixed effects and a control for the timing of the review. The findings reveal a similar pattern of results to the univariate results. We do not find any evidence that reviewers who wrote a negative review without a confirmed purchase are more upset with the firm than reviewers who wrote a negative review with a confirmed purchase.

Table 9
UPSET CUSTOMERS: SUBSEQUENT ORDERS ANALYSIS

	Any Review Without a Confirmed Transaction		*Only Reviews with a Confirmed Transaction*		*Difference*	
Years until next order	.2682	(.0117)	.2879	(.0037)	–.0197*	(.0119)
Purchase intervals until next order	1.0090	(.0652)	1.0853	(.0278)	–.0763	(.0881)
No subsequent order	16.37%	(.93%)	18.25%	(.31%)	–1.87%*	(1.01%)
No order in next purchase interval	34.58%	(1.33%)	38.48%	(.44%)	–3.90%***	(1.42%)
No order in next year	14.92%	(1.08%)	17.60%	(.37%)	–2.69%**	(1.21%)
More orders in next year versus prior year	34.25%	(1.44%)	25.96%	(.43%)	8.29%***	(1.41%)
More orders in next year versus prior average	59.30%	(1.40%)	53.98%	(.48%)	5.32%***	(1.59%)
Sample Sizes						
Years until next order	1,328		12,551			
Purchase intervals until next order	1,328		12,551			
No subsequent order	1,588		15,352			
No order in next purchase interval	1,284		12,533			
No order in next year	1,086		10,350			
More orders in next year versus prior year	1,086		10,350			
More orders in next year versus prior average	1,086		10,350			

*$p < .10$.
**$p < .05$.
***$p < .01$.
Notes: The unit of analysis is a reviewer × review date. We use observations that include at least one review with a rating equal to 1 (we report findings for all observations in the Web Appendix). The sample size changes because we restrict attention to observations for which we observe a complete postperiod. The sample size is also smaller when measuring the time or interval until the next order because we only consider observations for which there is a subsequent order.

products. They will even do so on products they have not purchased.[15]

Why would self-appointed brand managers be more likely to write a negative review? The French have a phrase that may help to answer this question: *Qui aime bien châtie bien*, which translates (approximately) to "Your best friends are your hardest critics." We investigated whether there is a relationship between the number of items that customers have purchased and the reviewers' product ratings. The pairwise correlation between a reviewer's average product rating and the number of items purchased is –.048 ($p < .01$). In other words, the best customers are the most negative reviewers.

We might also wonder why customers acting as self-appointed brand managers would write a review about a product they have not purchased given that they could write about the many products they have purchased. Perhaps these customers see a product for which they want to give feedback while browsing the firm's website. The urge to give feedback is prompted by what the reviewers see on the website rather than by a prior purchase, and the product review mechanism provides a convenient mechanism for them to do so.

We can investigate this explanation by asking the following question: When would a self-appointed brand manager be most likely to write a review? It may be that customers are more likely to react when they see a product that they did not expect. For example, if a customer who has only purchased women's apparel from the firm browses the firm's website and notices that the firm now sells pet products, this may prompt the self-appointed brand managers to provide feedback by clicking the button inviting a review.[16] We investigate this possibility by calculating the following measures:

- Prior units index: The total number of units of this item sold by the firm in the year before the date of the review. At the request of the retailer, we index this measure by setting the average to 100% for the reviews with a confirmed transaction.
- Niche items: Equal to 1 if prior units is in the bottom 10% of items with reviews and 0 if otherwise.
- Very niche items: Equal to 1 if prior units is in the bottom 1% of items with reviews and 0 if otherwise.
- Product age: Number of years between the date of the review and the date the item was first sold.
- New item: Equal to 1 if product age is less than two years and 0 if otherwise.
- New category: Equal to 1 if the maximum product age in the product category is less than two years and 0 if otherwise.

In Table 10, we report the average of each measure for reviews with and without confirmed transactions.[17] The findings reveal large (and highly significant) differences on all these measures. Reviews without a confirmed transaction are more likely to be written for items that were introduced recently. They also tend to be written for niche items with relatively small sales volumes. These findings are consistent with the explanation that customers are more likely to provide feedback to the firm when they see unexpected products on the firm's website.

In the Web Appendix, we report the rating distribution for different groupings of items. As we would expect, older products (that have survived longer) have higher ratings.

[15]A similar argument could also explain why community members contribute to building or zoning decisions in their community, even when those decisions do not directly affect the community members. In local hearings about variances for building permits, it is not unusual to receive submissions from community members who are not directly affected by the proposal. Like the review process, these hearings provide one of the most accessible mechanisms through which the community members can exert influence.

[16]In a related example, Harley-Davidson's introduction of a line of perfume ("Destiny by Harley-Davidson") reportedly prompted substantial negative feedback from its traditional customers (Haig 2003).

[17]In the Web Appendix, we control for valence by reporting the findings separately for reviews at each rating level.

Table 10

NICHE PRODUCTS AND NEW PRODUCTS

	Without a Confirmed Transaction	*With a Confirmed Transaction*	*Difference*
Prior units index	69.63%	100.00%	–30.37%* (1.14%)
Niche items	24.44%	9.26%	15.18%* (.24%)
Very niche items	8.57%	.61%	7.95%* (.08%)
Product age (years)	3.86	4.75	–.89%* (.04%)
New item	50.79%	44.03%	6.75%* (.41%)
New category	1.54%	1.15%	.39%* (.09%)

*$p < .01$.

Notes: The table reports averages for each measure separately for the samples of reviews with and without confirmed transactions. The sample sizes are 15,759 (reviews without a confirmed transaction) and 310,110 (reviews with a confirmed transaction). Standard errors are in parentheses.

Moreover, items that have higher sales volumes tend to have higher ratings. Because items without confirmed transactions are more likely to be niche or new products, this could contribute to the low rating effect. However, in our multivariate analysis of the product ratings, we replicate the low rating effect when including explicit controls for product age and product sales volumes (we also report a model with fixed-item effects). In addition, we replicate the low rating effect in our univariate results both in our within-item analysis and when comparing the rating distribution within different product age groups and different product sales volume quartiles. The low rating effect cannot be due to mere product differences.

Social Status

A third explanation is that reviewers are simply writing reviews to enhance their social status.[18] This explanation is related to a more general question: Why do customers ever write reviews with or without confirmed transactions? In an attempt to answer this question, some researchers have argued that customers are motivated by self-enhancement. Self-enhancement is defined as a tendency to favor experiences that bolster self-image and is recognized as one of the most important social motivations (Fiske 2001; Sedikides 1993). Wojnicki and Godes (2008) present empirical support that self-enhancement may motivate some customers to generate word of mouth (including reviews). Using both experimental and field data, they demonstrate that consumers "are not simply communicating marketplace information, but also sharing something about themselves as individuals" (Wojnicki and Godes 2008, p. 1). Other researchers, including Feick and Price (1987) and Gatignon and Robertson (1986), have proposed similar arguments.

Unlike some other websites, the retailer that provided data for this study does not celebrate its most prolific reviewers with titles such as "Elite Reviewers" (Yelp.com) or "Top Reviewer" (Amazon). However, it does identify reviewers by their chosen pseudonyms. Moreover, reviewers writing reviews without confirmed transactions do tend to be more prolific than other reviewers (see Table 8).

Self-enhancement may explain why reviewers write reviews for items they have not purchased. However, it does not immediately explain why these reviews are more likely to be negative. One possibility is that customers believe that they will be more credible if they contribute some negative reviews. This is consistent with research showing that readers perceive more negative reviewers to be more intelligent, competent, and expert than positive reviewers (Amabile 1983). These findings have been interpreted as evidence that reviewers who want to be perceived as more expert will contribute more negative opinions (Moe and Schweidel 2012; Schlosser 2005). In related research, Cheema and Kaikati (2010) show that people who have a high "need for uniqueness" are less willing to make positive recommendations about a product.

A further limitation of this explanation is that it does not directly explain why customers write reviews about products they have not purchased. Recall from Table 8 that, on average, these customers write approximately three reviews but have purchased 156 items. It is not clear why they do not enhance their status by writing a review about one of the many items they have purchased.

Distinguishing the "Self-Appointed Brand Manager" and "Social Status" Explanations

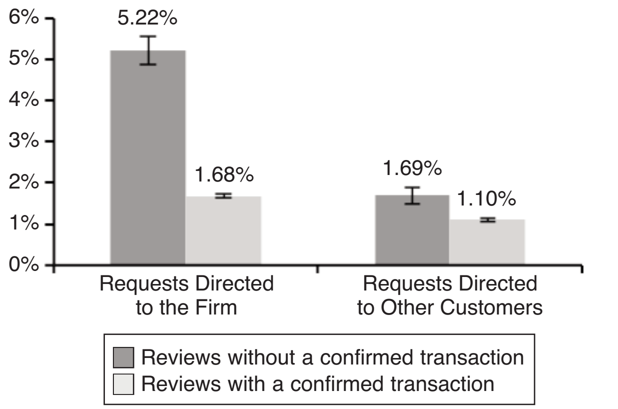
There is a subtle difference between the self-appointed brand manager and social status explanations in terms of with *whom* the reviewer is communicating. The self-appointed brand manager explanation anticipates that customers are providing feedback to the retailer. In contrast, under the social status explanation, reviewers are more likely to be providing advice to other customers. This distinction suggests an opportunity to differentiate the two explanations. We used text analysis to distinguish reviews that directed requests to the firm or offered advice to other customers.[19]

In Figure 3, we summarize the percentage of reviews with and without confirmed transactions that included either type of expression. The findings reveal that reviews without confirmed transactions are more than three times more likely to include requests directed at the company, consistent with these reviewers acting as self-appointed brand managers. Reviews without a confirmed transaction are also more likely to include advice directed to other customers, which is what we would expect if reviewers are trying to enhance their social status. However, although the findings offer support for both explanations, there is a clear difference in the relative magnitudes of the effects. This differ-

[18]For example, Harriet Klausner, a reviewer at Amazon, has contributed more than 25,000 book reviews (all reportedly unpaid), at a rate of approximately seven a day for a period of more than ten years. Notably, when queried about Mrs. Klausner and other examples of unpaid reviewers who acknowledged writing reviews for books they had not read, an Amazon spokesperson simply responded: "We do not require people to have experienced the product in order to write a review" (Streitfeld 2012).

[19]The text strings used to identify reviews that directed requests to the firm included "please," "bring back," "offer more," "carry more," and "go back to." The text strings used to identify reviews that offered advice to other customers included "if you are looking," "if you need," "if you want," "if you like," "if you order," "if you own," "if you buy," "if you purchase," "if you wear," and "if you prefer." To evaluate the recall and precision of this analysis, we randomly selected 50 reviews that the text analysis identified as reviews directing requests to the firm and 50 reviews identified as offering advice to other customers. We then asked a coder to read all 100 reviews and indicate whether the review was directed to either the customer or the firm. The recall and precision were 95% and 84% for the "directed to the firm" text strings and 100% and 84% for the "advice to other customers" text strings (see the Web Appendix).

Figure 3
EXPRESSIONS DIRECTED TO THE FIRM VERSUS TO OTHER CUSTOMERS



Notes: The figure reports the percentage of reviews that included each type of expression. The sample sizes include 15,759 reviews without and 310,110 reviews with confirmed transactions. The error bars are 95% confidence intervals. For detailed results, see the Web Appendix.

ence suggests that the self-appointed brand manager explanation plays a more prominent role in explaining why customers write reviews without confirmed transactions.

We caution that the text strings used to identify firm requests and customer advice are not the only expressions that reviewers could use for these purposes. For this reason, we should not conclude (for example) that only 5.22% of reviews without confirmed transactions included a request directed at the company. Instead, these expressions are cues that we use to measure the *relative* frequency of these requests or this advice.[20]

Summary

We present initial evidence that suggests that some reviewers who write reviews without confirmed transactions may be acting as self-appointed brand managers. We also present evidence that customers who write negative reviews without a confirmed purchase are no more upset with the firm than the customers who write negative reviews with a confirmed transaction. However, as we acknowledged at the beginning of this section, this evidence should be interpreted as an initial investigation of these explanations. Other explanations are also possible, and we hope that these findings encourage other authors to explore the phenomenon. In our final set of analyses, we investigate the implications of the low rating effect by examining whether it affects customers' purchases and the firm's revenue.

IMPLICATIONS FOR CUSTOMER PURCHASING BEHAVIOR AND FIRM REVENUE

To investigate whether the low rating effect has any impact on either the firm or its customers, we compare items' sales before and after the date of the review. Specifically, we calculate the change in the item's revenue for the year before versus the year after the review date. We then compare this change in revenue on reviews with a rating of 5 versus reviews with lower ratings. This is essentially a difference-in-difference approach (Bertrand, Duflo, and Mullainathan 2004) comparing the difference in revenue for reviews with different ratings. We are interested in whether a lower product rating is associated with a smaller increase (or larger decrease) in revenue earned. Note that the comparison of pre- versus postreview revenue controls for variation in revenue across items (some items sell more than others). Moreover, we do not assume that sales in the absence of the reviews would have been the same in the pre- and postperiods. Instead, the identifying assumption is that in the absence of the reviews, the expected change (pre- vs. post-) would have been the same.

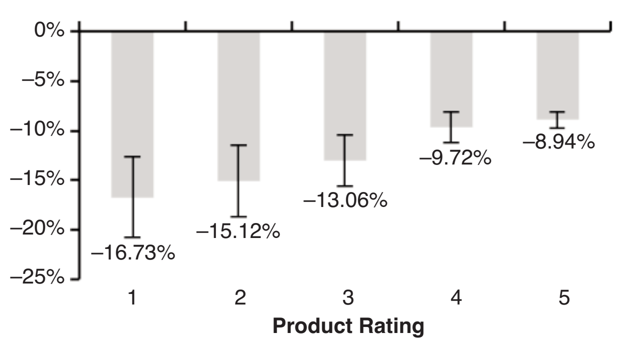
In Figure 4, we report the change in revenue between the one-year pre- and postperiods for each of the five rating levels. To ensure that we do not introduce any asymmetry in the magnitude of increases and decreases, we calculated the change in revenue as a percentage of the midpoint of the pre- and postperiod outcomes. The one-year periods control for seasonality, and we omit any item that was introduced or discontinued within these time windows. The unit of observation is an item × review date.[21] Because customers do not know whether a review is a confirmed transaction, we included all of the reviews in this analysis.[22] The findings

[20]In the Web Appendix, we repeat the analysis using a within-item approach (we also use a within-reviewer approach). We obtain the same pattern of findings, which rules out the possibility that the findings in Figure 3 can be explained by mere item differences.

[21]Recall that in our upset customer analysis (Table 9), the unit of analysis is a reviewer × review date (rather than an item × review date). When there are multiple reviews without confirmed transactions for the same item on the same day, we use the average of their product ratings.

[22]Although we have documented differences in the review text, there is an extensive literature stream documenting that humans are very poor at using these cues to detect deception (see, e.g., DePaulo 1994; Frank and Feeley 2003).

Figure 4
CHANGE IN ONE-YEAR REVENUE: REVIEWS WITHOUT CONFIRMED TRANSACTIONS



Notes: The figure reports the average change in revenue between one-year pre- and postperiods. The unit of analysis is an item × review date. We restrict attention to reviews written at least one year after the item was introduced and one year before the end of the data period. We also restrict attention to items with at least $1,000 in annual revenue. We calculated the percentage change using the average of the before and after revenues to ensure that increases and decreases are treated symmetrically. When there are multiple reviews without confirmed transactions for the same item on the same day, we use the average of their product ratings. Observations with a product rating equal to x include all reviewers for which the average rating is equal to x ± .5. The error bars are 95% confidence intervals.

reveal a consistent monotonic relationship. When the rating is more positive, there is a smaller decrease (or a larger increase) in revenue in the postperiod.

In the Web Appendix, we also report the findings when using units purchased (instead of revenue) and when weighting the observations by the number of reviews for that item that day. This weighting arguably provides a better measure of the average impact of an individual review. Finally, we also report the findings when using OLS to estimate the following model:

$$\ln(\text{Revenue}_{it}) = \alpha + \beta_1\text{Postperiod} + \beta_2\text{Rating\_1} + \beta_3\text{Postperiod} \times \text{Rating\_1} + \beta\mathbf{X} + \varepsilon.$$

The model includes two observations for each item × review date (i), including one observation for both the preperiod and the postperiod. In this first version of the model, we only include observations in which the average rating (for that item on that date) was either 1 or 5. The dependent variable measures the log of Revenue in that period, Postperiod is a binary variable identifying whether the observation is for the postperiod, and Rating_1 is a binary variable identifying whether the rating was 1 (vs. 5). The other control variables include fixed item effects, the date of the review (measured in years after the date of the firm's first review), the number of previous reviews of that item, and the average rating on the previous reviews. Because the average rating on previous reviews is only well-defined if there is at least one previous review, when there are no previous reviews, we set this average rating to 0 and include a binary variable identifying these observations.

This is a classic difference-in-difference specification, in which the reviews with a rating of 5 represent the control. The coefficient of interest is β_3, which measures whether the change in revenue between the preperiod and postperiod is higher or lower if there is a rating of 1 versus a rating of 5. We report the findings in the Web Appendix, in which we cluster the standard errors at the item level. We also report a version of the model using all of the rating levels (reviews with a rating of 5 again represent the control) together with models in which we weight the observations by the number of reviews for that item that day. All of these robustness checks yield a similar pattern of results, replicating the univariate findings. As might be expected, the results are stronger when we weight the observations.

It is possible that the positive relationship between the product rating and the change in revenue reflect the reviewers' predictive abilities. However, the difference-in-difference nature of the analysis makes this explanation unlikely. Although it is plausible that reviewers can predict which items will earn less revenue, the findings measure the change in revenue rather than the base level of revenue. It is less clear why reviewers would be able to predict the change in revenue. An alternative interpretation is that the reviews influence future sales performance. This is consistent with mounting evidence elsewhere in the literature stream that reviews can affect product sales (see, e.g., Chevalier and Mayzlin 2006). This second interpretation suggests that the low rating effect may have important implications for the firm and its customers. In particular, the disproportionate number of low ratings may dissuade customers from buying products they would otherwise purchase.

We can estimate the potential impact of the low rating effect on firm sales by calculating the average change in sales if the distribution of product ratings were the same for reviews without confirmed transactions as for reviews with confirmed transactions. For each review without a confirmed transaction, we estimate (using the one-year comparison) that revenue is lowered by approximately .56% compared with the previous year's revenue. Items that have reviews without confirmed transactions have 3.93 of these reviews on average, and so the aggregate impact of the low ratings on these items is a reduction in revenue by approximately 2.2%. We caution that this estimate is best interpreted as an upper bound because it ignores any substitution of this revenue to other products.

CONCLUSIONS

We study customer reviews of private label products sold by a prominent apparel retailer. Our analysis compares the product ratings on reviews for which we observe that the customer has a confirmed transaction for the product with reviews that lack confirmed transactions. The findings reveal that the 5% of reviews for which there is no observed confirmed transaction have significantly lower product ratings than the reviews with confirmed transactions. There are also significant differences in the content of the text comments.

Reviews without confirmed transactions are contributed by 12,474 individual customers. The low rating effect is particularly prominent among the 11,944 customers who submitted only one or two reviews without confirmed transactions. They are some of the firm's most valuable customers, who on average have each purchased more than 100 products. The number of reviewers and the frequency of their purchases make it unlikely that the phenomenon can be attributed to competitors. The low rating effect seems to be due to actual customers engaging in this behavior for their own intrinsic interests. In this respect, the findings represent evidence that the manipulation of product reviews is not limited to strategic behavior by competing firms. Instead, the phenomenon may be far more prevalent than previously thought.

We are able to rule out several alternative explanations for the low rating effect. The effect cannot be attributed to item differences, reviewer differences, gift recipients, purchases by other customers in the household, customers misidentifying items, changes in item numbers, purchases on secondary markets, unobserved transactions (in retail stores), complaints about non-product-related issues (shipping or service complaints), or differences in the timing of the reviews. We caution that despite this long list of robustness checks, we cannot conclusively establish that customers never purchased the item—only that we can find no record of a purchase. However, any alternative explanation must explain not only why we do not observe a purchase but also why these reviews have low ratings and why there are significant differences in the review text.

A second limitation of the study involves the absence of direct evidence of deception. This limitation is common to almost all studies of deception that do not rely on constructed stimuli. As with other studies of deception in online reviews, we infer deception from behavioral patterns that deviate from behavior that is thought to be truthful. We rely on two sources of evidence: First, we show that reviews without confirmed transactions are more likely to contain linguistic cues associated with deception. Second, we replicate the findings using a sample of reviewers who self-identified that they purchased the items. However, we emphasize that our results should not be interpreted as evidence that all the reviews without confirmed transactions are deceptive.

This article has several important managerial implications. Expedia.com's model of only allowing customers who have purchased the product to write a review is one approach to resolving the phenomenon that we document. The firm that participated in this study could adopt a similar policy, one that only allows reviewers to submit reviews for items that they have purchased. Alternatively, the firm could follow Amazon's policy of identifying whether a review matches a confirmed transaction. If customers become aware that the phenomenon is as widespread as the findings in this article suggest, conditioning the acceptance of reviews on a prior purchase may become the industry standard. This has another important implication. If, in the long run, reviews at a website are only considered credible when they are linked to a purchase, this may harm the business model of firms that report reviews that are not linked to transactions. For example, these findings may raise concerns about the current business models of firms such as Yelp.com and TripAdvisor.com. In the future, these firms may form relationships with partners that can provide access to transaction information.

As we discussed in the beginning of this article, reviewers represent the extreme tail of all customers. Although their preferences might not be representative of other customers, their reviews do influence the purchasing decisions of other customers. This raises important questions about whether (or when) reviews are accretive to social welfare. The nonrepresentative nature of reviews may also have implications for competition. If firms all respond by designing products or setting prices to target a small group of reviewers, they may forgo the opportunity to differentiate (see, e.g., Simester 2011).

Further research could evaluate how the level of deception varies across reviewers or product categories. Although not all researchers will have access to the type of data provided by the apparel retailer that participated in this study, all researchers do have access to data from Amazon and similar sites. The replication of our findings using the book reviews at Amazon may facilitate further research of this type by validating the use of the Amazon Verified Purchase cue as an indicator of deception.

APPENDIX: RULING OUT ALTERNATIVE EXPLANATIONS

Could the Low Ratings Be Due to Item Differences?

It is possible that the reviews without confirmed transactions are written for products that are different (and of lower quality) than the reviews with confirmed transactions. To investigate this possibility, we conducted a within-item comparison using the 3,779 items for which we have reviews with and without confirmed transactions. For each item, we separately calculated the mean rating and the frequency of each rating level for reviews with and without confirmed transactions. We then calculated the difference in these measures and average the differences across all 3,779 items. We report the findings in the Web Appendix (in which we also include a more complete description of this analysis). They closely match the findings in Table 1.

To reinforce this finding, we also replicated the ratings comparison separately using each of the ten largest product categories and when grouping the products according to their product ages and sales volumes. Finally, we also estimated an OLS model with fixed item effects. The low rating effect survives all of these robustness checks (we report the findings in the Web Appendix). We conclude that the difference in the ratings between reviews with and without confirmed transactions cannot be attributed to mere item differences.

Could the Low Ratings Be Due to Reviewer Differences?

It is possible that the reviewers who wrote reviews for which we have no confirmed transactions are different (and more negative) than reviewers who wrote reviews for which we do have confirmed transactions. We investigated this possibility using a similar approach to the item differences analysis. Specifically, we compared the ratings in which the same reviewer had written some reviews with a confirmed transaction and some reviews without a confirmed transaction. For each of these 5,234 reviewers, we separately calculated the mean rating and the frequency of each rating level for reviews with and without confirmed transactions. We then calculated the difference in these measures for each reviewer and averaged these differences across the 5,234 reviewers. We report the findings in the Web Appendix.[23]

This within-reviewer comparison again reveals the same pattern of results. Reviews without confirmed transactions tend to be more negative than reviews with confirmed transactions, even though the same reviewers wrote both sets of reviews. We conclude that the difference cannot be attributed to reviewer differences. These findings also provide an initial indication that the effect is not limited to a handful of rogue reviewers. Instead, it seems that the effect extends across several thousand reviewers.

It is possible that the customers have purchased the items, but we are unable to match their transactions with their reviews. We investigate this possibility next by examining limitations in our data and/or errors by the customers that could lead us to incorrectly overlook a customer's prior purchase.

Could Customers Have Purchased the Items on a Secondary Market?

Although the initial sale of the firm's products always occurs through one of the firm's retail channels, the items may be resold on secondary markets, such as eBay and Craigslist. Because the items are relatively low priced and the firm offers a very generous return policy, the firm believes that there is relatively little trade in its products on secondary markets. A search for the company's products on eBay revealed approximately 15,000 units available for sale. Although this may suggest a substantial volume of trade, it appears negligible when compared with the total volume of sales through the firm's retail channels.

[23]Because many customers write only one review without a confirmed transaction, in the Web Appendix we report findings for reviewers who have at least three reviews without a confirmed transaction (and at least one review with a confirmed transaction).

We used two approaches to investigate whether the reviews without confirmed transactions could have been contributed by customers purchasing from a secondary market. First, we searched the review text for the strings "ebay" and "craigslist" (the search was not case sensitive). We found only 2 reviews (out of the 325,869) in which the reviewer identified that he or she had purchased the item through eBay, and no instances in which he or she had purchased the item through Craigslist. Although we would not expect all the customers who purchased through a secondary market to report that they had done so, it is notable that essentially no reviewers did so.

Second, one category that we might expect customers to be reluctant to purchase on a secondary market is "underwear." A detailed inspection of the eBay product listings (which are grouped by product category) confirmed that none of 15,000 of the company's items available on eBay are in the underwear category. In comparison, 3,200 of the product reviews are for underwear items, suggesting that underwear is a category in which we can repeat our analysis with confidence that the outcome is unaffected by sales in secondary markets. The Web Appendix reports the findings. Although the reduction in the sample size reduces the statistical significance of the results, we continue to observe the same pattern of results reported previously. In particular, there are twice as many ratings of 1 when there is no confirmed transaction condition compared with when there is a confirmed transaction. We conclude that purchases on secondary markets cannot be the only explanation for the low rating effect.

Complaints About Shipping or Customer Service

The product review mechanism is specific to a product and is designed for customers to provide feedback about that product. However, it is possible that a customer may provide feedback about topics that are not directly related to a product, such as the firm's shipping policies or customer service. As we discussed previously, the firm offers other channels for customers to provide feedback that is not directly related to a specific product. The firm's website invites customers to submit feedback by telephone, e-mail, blog, a story-sharing site, and several social media sites hosted by the firm (including Facebook, Twitter, Foursquare, and Google+). Despite the availability of these other channels, it is possible that customers use the review mechanism to provide feedback about general issues rather than specific products. This could explain why reviewers write reviews without having purchased the item and could also explain why these reviews tend to be more negative.

To investigate this possibility, we searched the review text to identify reviews in which customers provided feedback about either customer service or shipping policies. To identify customer service feedback, we searched for the words "service" or "rep." For shipping policy feedback, we searched for "shipping," "postage," and "charges." The recall and precision for both sets of text strings are 100% (see the Web Appendix). Inspection of the reviews that contained these words indicated that they almost always included some feedback related to these issues. However, with very few exceptions, the primary focus of the review was the product itself. We found almost no reviews that focused solely on customer service or shipping policies without also addressing a product-related issue.

If the reviews without confirmed transactions result from customers using the product review process to provide feedback about customer service or shipping policies, they should be more likely to mention these words. Therefore, we compared the presence of these words in reviews with and without confirmed transactions. The findings (reported in the Web Appendix) indicate that reviewers are actually significantly *less* likely to make comments about shipping policies when writing reviews without confirmed transactions. Moreover, there is essentially no difference in the frequency of comments about customer service. We conclude that the reviews without confirmed transactions do not seem to be explained by customers using the review mechanism to provide feedback about firm policies that are unrelated to specific products.

Could the Low Ratings Be Due to Customers Misidentifying Items?

One reason we may have overlooked a confirmed transaction is that customers could have incorrectly identified the item number. Recall that reviews are submitted by clicking on a button on the product page for each item. It is possible that some customers purchase an item and mistakenly submit a review for a similar but different item.

A closely related explanation is that customers may write reviews for different versions of the same product. When the firm updates the design of an item, it will sometimes assign a new item number to the updated product. In our analysis, we identified products at a relatively aggregate level so that all sizes and colors are included under the same item number. This ensures that reviews without confirmed transactions cannot be attributed to customers misidentifying the color or size of the item. However, it is possible that reviewers have purchased an earlier version of an item with a different item number than the item they reviewed.

To investigate these possibilities, we used an even broader level of aggregation to match reviews with the reviewers' purchases. In particular, we repeated our analysis when identifying items at the product subcategory level. Examples of subcategories include "women's gingham shirts" and "men's chino shorts." The items with reviews are distributed across 3,655 subcategories. On the one hand, the advantage of using this subcategory level of aggregation is that it essentially excludes the possibility that a confirmed transaction is overlooked because either customers misidentify another item in the subcategory or the item number has changed. On the other hand, this approach increases the probability that we incorrectly identify a review as having a prior purchase, when the customer's prior purchases in the subcategory were for completely different items.

When using subcategories to identify items without confirmed transactions, we omitted 115 reviews for items not associated with a subcategory. Of the remaining 325,754 reviews, we found 9,150 reviews (2.81%) without a confirmed transaction. This reduction in the percentage of reviews without a confirmed transaction reflects the broader definition of an "item" when matching at the subcategory level. In the Web Appendix, we report the distribution of product ratings for reviews with and without confirmed transactions using this subcategory approach. The pattern of findings is essentially identical to those reported in Table 1. We conclude that the low rating effect cannot be explained by misidentified items or customers writing reviews on later versions of items that they had previously purchased.

Could the Low Ratings Be Due to Unobserved Transactions in the Retail Stores?

When making purchases in the firm's retail stores, almost all customers use a credit card. This makes it relatively easy for the firm to associate the customer with a unique account number in its transaction database. However, on the (rare) occasions that a customer pays cash for a purchase in a retail store, there may be too little information to identify the customer. This could result in customers writing a review for an item that they have purchased, but we never observe the transaction. Note that this almost never occurs when customers purchase through the catalog or Internet channels, because they provide more identifying personal information to the firm when purchasing in these channels.

To explain the low rating effect, unobserved transactions in retail stores must yield lower product ratings. We can investigate whether transactions in retail stores typically have lower ratings by inspecting the reviews for which we do have confirmed transactions. In the Web Appendix, we report the distribution of product ratings according to the retail channel in which the purchase occurred.[24] Our findings show that product ratings are highest when the confirmed transaction occurred in a retail store. A simple explanation is that retail stores typically offer customers the best opportunity to inspect items before they purchase. Higher ratings on items purchased in retail stores suggest that if the reviews without confirmed transactions were unobserved purchases in retail stores, we would expect higher (not lower) ratings on these reviews. Moreover, in general, the differences in the ratings across the three retail channels are small, making it unlikely that the low ratings for reviews without a transaction are due to customers making unobserved purchases from a specific retail channel.

We can further investigate whether the low rating effect results from unobserved purchases in retail stores by identifying customers who are unlikely to purchase in one of this retailer's stores. We do so in two ways. First, we use the customers' individual purchase histories to exclude customers who ever purchased in one of the firm's retail stores. Reviewers in our data set have each purchased an average of more than 100 items, so this is a relatively strong filter. Second, we use the customers' zip codes to exclude any customer who lives within 400 miles of a retail store. In the Web Appendix, we compare the average ratings for the remaining reviewers. The pattern of findings almost perfectly replicates the findings in Table 1. In particular, the average rating and the percentage of (low) ratings equal to 1 is essentially unchanged.

We can also use variation across items to investigate the retail store explanation. We searched for a sample of items that are only available for purchase through the firm's catalog or Internet sites and are not available in its retail stores. Unfortunately, there are few items with zero retail store transactions because the firm typically offers at least one color or size variant of each item in its stores. However, there are items that have very few retail store transactions. In particular, we focused on items for which more than 98% of all purchases occurred through the catalog and Internet channels (less than 2% occurred in retail stores).[25] Notably, there is a slightly higher proportion of reviews without confirmed transactions in this restricted sample (7.4%) compared with the complete sample (4.8%), which is not what we would expect if these reviews reflect purchases in retail stores. We then repeated our analysis when restricting attention to these items. We again observed the same pattern of findings.

Finally, in the Web Appendix, we investigate whether the product ratings are lower for items for which a larger percentage of sales occur in retail stores (vs. the catalog or Internet channels). The proportion of negative ratings is actually significantly negatively correlated with the proportion of items sold in retail stores. In other words, items with a higher proportion of sales in retail stores tend to have more positive ratings. Moreover, the difference in ratings between reviews with and without a prior transaction is very stable and seemingly not affected by what proportion of an item's sales occur in retail stores.[26] We conclude that the low rating effect is unlikely to be explained by customers making unobserved purchases at retail stores.

Could the Low Ratings Be Due to Differences in the Timing of the Reviews?

Our data record the date that each review was written. A comparison of these dates reveals that, on average, reviews without confirmed transactions were written slightly earlier than reviews with confirmed transactions. The average review date is approximately 3.5 months earlier for reviews without confirmed transactions. To investigate whether these timing differences could have contributed to the lower product ratings, we calculated the average ratings for the two sets of reviews in each year. These average ratings are reported in the Web Appendix.

For both sets of reviews, we observe that reviews written later actually have *lower* average ratings, consistent with previous research that reviews have become more negative over time (Godes and Silva 2012; Li and Hitt 2008; Moe and Trusov 2011). However, it is the opposite of what we would expect if the low rating effect was due to timing differences. To further investigate this explanation, we also estimated an OLS model with fixed effects to control for the day the review was created (we report these findings in the Web Appendix). The low rating effect survived and was actually strengthened by these controls for the timing of the review.

We also investigated another timing-related explanation: If a transaction occurred a long time in the past, there may

[24]We omitted a handful of reviews for which the customer purchased the item in multiple channels before writing the review. For example, a customer may have purchased a pair of pants in a retail store and another pair of the same pants (in a different transaction) over the Internet.

[25]This restriction results in a sample of reviews in which on average less than 1% (.87%) of all transactions for those items occur in retail stores. We use all of the transactions (by any customer) when calculating how many purchases occurred in retail stores.

[26]In our multivariate analysis replicating the low rating effect, we include explicit controls for the percentage of units (of that item) that are sold in retail stores.

be a higher likelihood of errors in matching a customer's transaction with the customer's review. It is also possible that there are more low ratings when the transaction occurred a long time before the review date. To investigate this explanation, we used the sample of reviews that have confirmed transactions. This revealed that when there is a longer interval between the date of the transaction and the date of the review, the reviews are slightly *less* likely to have low ratings. We conclude that the low rating effect does not seem to result from transactions occurring a long time before the review date.[27]

Finally, in the Web Appendix we also report findings when we grouped the items on the basis of the age of the item at the date of the review: less than one year, one to two years, two to four years, four to six years, six to ten years, or more than ten years. We then replicate our analysis separately on each of these groups of observations. The pattern of findings remains unchanged across all of these replications.

REFERENCES

Amabile, Teresa M. (1983), "Brilliant but Cruel: Perceptions of Negative Evaluators," *Journal of Experimental Social Psychology*, 19 (2), 146–56.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2004), "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 119 (1), 249–75.

Bond, Charles F. and Bella M. DePaulo (2006), "Accuracy of Deception Judgments," *Personality and Social Psychology Review*, 10 (3), 214–34.

Burgoon, Judee K., J.P. Blair, Tiantian Qin, and Jay F. Nunamaker Jr. (2003), "Detecting Deception Through Linguistic Analysis," in *Proceedings of the 1st NSF/NIJ Conference on Intelligence and Security Informatics*, Hsinchun Chen, Richard Miranda, Daniel D. Zeng, Chris Demchak, Jenny Schroeder, and Therani Madhusudan, eds. Berlin: Springer-Verlag, 91–101.

Cheema, Amar and Andrew M. Kaikati (2010), "The Effect of Need for Uniqueness on Word of Mouth," *Journal of Marketing Research*, 47 (June), 553–63.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (August), 345–54.

Clark, Patrick (2013), "New York State Cracks Down on Fake Online Reviews," *BusinessWeek*, (September 23), (accessed February 24, 2014), [available at http://www.businessweek.com/articles/2013-09-23/new-york-state-cracks-down-on-fake-online-reviews].

Dellarocas, Chrysanthos (2006), "Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms," *Management Science*, 52 (20), 1577–93.

DePaulo, Bella M. (1994), "Spotting Lies: Can Humans Learn to Do Better?" *Current Directions in Psychological Science*, 3 (3), 83–86.

———, James J. Lindsay, Brian E. Malone, Laura Muhlenbruck, Kelly Charlton, and Harris Cooper (2003), "Cues to Deception," *Psychological Bulletin*, 129 (1), 74–118.

Donath, Judith (1999), "Identity and Deception in the Virtual Community," in *Communities in Cyberspace*, Marc A. Smith and Peter Kollock, eds. New York: Routledge, 29–59.

Eisenberger, Robert, Patrick Lynch, Justin Aselage, and Stephanie Rohdieck (2004), "Who Takes the Most Revenge? Individual Differences in Negative Reciprocity Norm Endorsement," *Personality and Social Psychology Bulletin*, 30 (6), 787–99.

[27]In our multivariate analysis replicating the low rating effect, we include explicit controls for both the date of the review and the age of the item.

Feick, Lawrence and Linda L. Price (1987), "The Market Maven: A Diffuser of Marketplace Information," *Journal of Marketing*, 51 (January), 83–97.

Fiske, Susan T. (2001), "Five Core Social Motivates, Plus or Minus Five," in *Motivated Social Perception: The Ontario Symposium*, Vol. 9, Steven J. Spencer, Steven Fein, Mark P. Zanna, and James M. Olsen, eds. Mahwah, NJ: Lawrence Erlbaum Associates.

Frank, Mark G. and Thomas H. Feeley (2003), "To Catch a Liar: Challenges for Research in Lie Detection Training," *Journal of Applied Communication Research*, 31 (1), 58–75.

Gatignon, Hubert and Thomas S. Robertson (1986), "An Exchange Theory Model of Interpersonal Communication," in *Advances in Consumer Research*, Vol. 13, Richard J. Lutz, ed. Provo, UT: Association for Consumer Research, 534–38.

Godes, David and Dina Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23 (4), 545–60.

——— and ——— (2009) "Firm-Created Word-of-Mouth Communication: Evidence from a Field Study," *Marketing Science*, 28 (4), 721–39.

——— and Jose C. Silva (2012), "Sequential and Temporal Dynamics of Online Opinion," *Marketing Science*, 31 (3), 448–73.

Haig, Matt (2003), *Brand Failures: The Truth About the 100 Biggest Branding Mistakes of All Time*. Sterling, VA: Kogan Page Business Books.

Hsu, Chiao-Fang, Elham Khabiri, and James Caverlee (2009), "Ranking Comments on the Social Web," in *Proceedings of the 2009 International Conference on Computational Science and Engineering*, Vol. 4. Washington, DC: IEEE Computer Society, 90–97.

Humphreys, Sean L., Kevin C. Moffitt, Mary B. Burns, Judee K. Burgoon, and William F. Felix (2011), "Identification of Fraudulent Financial Statements Using Linguistic Credibility Analysis," *Decision Support Systems*, 50 (3), 585–94.

Jindal, Nitin and Bing Liu (2007), "Analyzing and Detecting Review Spam," in *Proceedings of the Seventh IEEE International Conference on Data Mining*, N. Ramakrishnan, O.R. Zaiane, Y. Shi, C.W. Clifton, and X.D. Wu, eds. Los Alamitos, CA: IEEE Computer Society, 547–52.

Kornish, Laura J. (2009), "Are User Reviews Systematically Manipulated? Evidence from Helpfulness Ratings," working paper, Leeds School of Business, University of Colorado Boulder.

Lee, Thomas Y. and Eric T. Bradlow (2011), "Automated Marketing Research Using Online Customer Reviews," *Journal of Marketing Research*, 48 (October), 881–94.

Li, Xinxin and Lorin M. Hitt (2008), "Self-Selection and Information Role of Online Product Reviews," *Information Systems Research*, 19 (4), 456–74.

Luca, Michael and Georgios Zervais (2013), "Fake It Till You Make It: Reputation, Competition, and Yelp Review Fraud," working paper, Harvard University.

Mayzlin, Dina (2006), "Promotional Chat on the Internet," *Marketing Science*, 25 (2), 155–63.

———, Yaniv Dover, and Judith Chevalier (2013), "Promotional Reviews: An Empirical Investigation of Online Review Manipulation," *American Economic Review*, forthcoming.

Moe, Wendy W. and David Schweidel (2012), "Online Product Opinions: Incidence, Evaluation, and Evolution," *Marketing Science*, 31 (3), 372–86.

——— and Michael Trusov (2011), "The Value of Social Dynamics in Online Product Ratings Forums," *Journal of Marketing Research*, 48 (June), 444–56.

Mukherjee, Arjun, Bing Liu, and Natalie Glance (2012), "Spotting Fake Reviewer Groups in Consumer Reviews," in *Proceedings of the 21st International Conference on World Wide Web*. New York: Association for Computing Machinery, 191–200.

Newman, Matthew L., James W. Pennebaker, Diane S. Berry, and Jane M. Richards (2003), "Lying Words: Predicting Deception from Linguistic Styles," *Personality and Social Psychology Bulletin*, 29 (5), 665–75.

O'Mahony, Michael P. and Barry Smyth (2009), "Learning to Recommend Helpful Hotel Reviews," in *Proceedings of the Third ACM Conference on Recommender Systems*. New York: Association for Computing Machinery, 305–308.

Ott, Myle, Yejin Choi, Claire Cardie, and Jeffrey T. Hancock (2011), "Finding Deceptive Opinion Spam by Any Stretch of the Imagination," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*. Portland, OR: Association for Computational Linguistics, 309–319.

Schlosser, Ann E. (2005), "Posting Versus Lurking: Communicating in a Multiple Audience Context," *Journal of Consumer Research*, 32 (2), 260–65.

Sedikides, Constantine (1993), "Assessment, Enhancement, and Verification Determinants of the Self-Evaluation Process," *Journal of Personality and Social Psychology*, 65 (2), 317–38.

Simester, Duncan I. (2011), "When You Shouldn't Listen to Your Critics," *Harvard Business Review*, (June), (accessed March 18, 2014), 42.

Streitfeld, David (2012), "Giving Mom's Book Five Stars? Amazon May Cull Your Review," *The New York Times*, (December 23), (accessed February 25, 2014), [http://www.nytimes.com/2012/12/23/technology/amazon-book-reviews-deleted-in-a-purge-aimed-at-manipulation.html].

Toma, Catalina L. and Jeffrey T. Hancock (2012), "What Lies Beneath: The Linguistic Traces of Deception in Online Dating Profiles," *Journal of Communication*, 62 (1), 78–97.

Wojnicki, Andrea C. and David B. Godes (2008), "Word-of-Mouth as Self-Enhancement," working paper, University of Toronto.

Wu, Guangyu, Derek Greene, Barry Smyth, and Padraig Cunningham (2010), "Distortion as a Validation Criterion in the Identification of Suspicious Reviews," in *Proceedings of the First Workshop on Social Media Analytics, Association for Computing Machinery*. New York, 10–13.

Yoo, Kyung-Hyan and Ulrike Gretzel (2009), "Comparison of Deceptive and Truthful Travel Reviews," in *Information and Communication Technologies in Tourism 2009: Proceedings of the International Conference*, Wolfram Hopken, Ulrike Gretzel, and Rob Law, eds. Vienna: Springer-Verlag, 37–47.

Zhou, Lina (2005), "An Empirical Investigation of Deception Behavior in Instant Messaging," *IEEE Transactions on Professional Communication*, 48 (2), 147–60.

———, Judee K. Burgoon, Jay F. Nunamaker Jr., and Douglas Twitchell (2004), "Automating Linguistic-Based Cues for Detecting Deception in Text-Based Asynchronous Computer-Mediated Communication," *Group Decision and Negotiation*, 13 (1), 81–106.

———, ———, and Douglas P. Twitchell (2003), "A Longitudinal Analysis of Language Behavior of Deception in E-mail," in *Proceedings of the 1st NSF/NIJ Conference on Intelligence and Security Informatics*, Hsinchun Chen, Richard Miranda, Daniel D. Zeng, Chris Demchak, Jenny Schroeder, and Therani Madhusudan, eds. Berlin: Springer-Verlag, 102–110.

Zuckerman, Miron and Robert E. Driver (1985), "Telling Lies: Verbal and Nonverbal Correlates of Deception," in *Multichannel Integrations of Nonverbal Behavior*, Aron W. Siegman and Stanrey Feldman, eds. Hillsdale, NJ: Lawrence Erlbaum Associates, 129–44.