Games and Economic Behavior 87 (2014) 554–570

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Why do people rate? Theory and evidence on online ratings ☆

Jonathan Lafky *

Lafayette College, Easton, PA 18042, United States

ARTICLE INFO

Article history:
Received 26 March 2011
Available online 5 March 2014

JEL classification:
C91
D64
D83
L86

Keywords:
Online ratings
Altruism
Punishment

ABSTRACT

The rapid growth of online retail in the last decade has led to widespread use of consumer-generated ratings. This paper theoretically and experimentally identifies influences that drive consumers to rate products and examines how those factors can create distortions in product ratings. By manipulating payoffs and effectively "deactivating" either the buyer or seller side of an artificial laboratory market, raters' behavior is decomposed into buyer-centric and seller-centric components. The cost of providing a rating also plays a major role in influencing rating behavior, with high and low quality sellers being rated more often than those of moderate quality.

1. Introduction

Internet commerce is a large and rapidly growing component of the economy. Internet retail accounted for $224.3 billion in sales for 2012, up 16.2% from 2011. Typical growth over the past decade has been even higher, averaging approximately 20% annually. Online retail's share of total U.S. retail has also increased tremendously over the past decade, climbing from just over 1% in 2001 to 5.2% by the end of 2012.[1]

The rapid growth and popularity of internet retail is not surprising. Virtually any good can be purchased on the internet, in every model, style, or color produced. The enormous selection offered to consumers means that they must often choose between several goods with similar observable characteristics but potentially different levels of quality. Without firsthand experience, it may be difficult or impossible for consumers to tell which of several similar-looking products is of the highest quality.

In an effort to alleviate this problem and to encourage sales, many internet retailers provide customer-based rating and review systems for their products. In these systems, consumers (sometimes restricted only to previous buyers) are allowed to post written reviews as well as numerical scores for products. These ratings are then made available to future buyers to inform them of the product's qualities, allowing them to make more informed purchases. For example, Amazon.com allows customers to leave ratings between one and five stars in one-star increments, as well as written comments about products they have purchased.

☆ This paper was previously titled "Why do people write reviews? Theory and evidence on online ratings."

* Fax: +1 (610) 330 5715.
E-mail address: lafkyj@lafayette.edu.

[1] U.S. Census Bureau, December 2013 Monthly Retail Trade Report. Retrieved from http://www.census.gov/retail/.

http://dx.doi.org/10.1016/j.geb.2014.02.008
0899-8256/

The average review score can vary considerably for products that have otherwise similar characteristics, and may be the only insight consumers have into a product's unobservable qualities before they buy. As Chevalier and Mayzlin (2006) demonstrate, ratings can significantly influence buyers' behavior and have a substantial impact on the success or failure of a product. But why are ratings given in the first place? Are people taking time to give these ratings in order to help their anonymous fellow shoppers, or are they writing out of gratitude or anger that they feel towards online merchants? Are raters equally likely to evaluate all products, or do they speak up only if they have a strong opinion? This paper examines possible motivations for the provision of numerical ratings in a theoretical framework and then isolates those motivations in an experimental setting.

To preview the results, I find evidence that consumers are motivated by concern for both buyers and sellers when they decide to rate products. Making rating less attractive through the introduction of a small cost has a large effect on the volume and distribution of ratings. Ratings in the presence of a cost take on a U-shaped distribution, which can lead to average ratings that are not representative of true quality. A possible solution to this problem is to provide small discounts to consumers who provide ratings, thereby compensating for any inconveniences or opportunity costs associated with rating products.

This paper focuses solely on numerical ratings and does not examine textual reviews. While consumers may ultimately be influenced by written reviews, they are not incorporated in the numerical score that is typically the first signal consumers receive as to a product's quality. Additionally, for the same reason that written reviews may be useful, they are also difficult to analyze in a rigorous and objective manner. Just as written reviews express nuances not easily captured through a numerical score, the reviews themselves are not easily quantified without imposing substantial subjectivity. Nonetheless, some of this paper's insights on consumer rating behavior may be applicable to written reviews as well. In particular it seems plausible that polarization may occur in written reviews for the same reasons that it occurs in numerical ratings. Extending these results to the domain of written reviews would be a useful area for future research.

The remainder of the paper is organized as follows. Section 2 surveys relevant past research. Section 3 provides motivating data and poses the basic questions to be addressed. Section 4 introduces a theoretical framework for analyzing rating behavior and isolating concern for sellers from concern for buyers. Section 5 lays out the experimental design and hypotheses. Section 6 presents results from the experiments while Section 7 discusses implications of those findings. Section 8 concludes.

2. Related literature

There is a small but growing literature on online ratings, with most existing work focusing on how consumers are influenced by ratings and how well those ratings can predict market outcomes. Much of the literature takes the existence of ratings as a given, avoiding the questions of why or how accurately the ratings are created. Chevalier and Mayzlin (2006), for example, convincingly show that ratings influence consumers' book purchasing behavior, but they do not examine whether the ratings accurately reflect book quality.

Other authors have questioned the value of ratings, arguing that they may predict future sales without actually influencing them. Duan et al. (2008) and Dellarocas et al. (2004) argue that consumer-generated ratings for movies are simply a gauge of underlying word-of-mouth communication, rather than a driver of movie success or failure. Such arguments over causality illustrate one of the major advantages of moving ratings research into the laboratory, where it is easier to cleanly identify causal relationships between rater incentives, ratings and purchases. Bolton et al. (2004) use such an experimental setting to show a causal relationship between feedback and trustworthiness of sellers in a simulated market. In their setting, however, feedback is automatic, and thus is not subject to any potential biases or strategic behavior that raters may exhibit in practice.

There is a small body of existing research on behaviors that may influence ratings. This work largely focuses on the potential for biases from self-selection, as in Li and Hitt (2008) and Hu et al. (2009). Self-selection across time is explored by Li and Hitt, who consider possible distortions in ratings for newly released products on Amazon.com. They find that products experience consistent rising and falling patterns across time, which can be explained by early adoption among "avid fans" and a later backlash from average consumers. Hu et al. use similar insights to explain the tendency for long-term ratings to take a J-shaped distribution. They propose two biases to explain the distribution: A "moan and groan" underreporting bias, and a self-selection "purchasing" bias, similar to that proposed by Li and Hitt. Li and Xiao (forthcoming) also find a bias in rating behavior, with consumers being more sensitive to the cost of rating for high quality sellers than for low quality sellers.

Self-selection is not the only issue relevant to rating generation, however, and Wang (2010) demonstrates that other factors such as social identity and anonymity can play major roles in consumers' decision to provide ratings. He finds that a strong sense of social identity considerably increases the quantity and quality of ratings. In a similar vein, Chen et al. (2010) use social comparisons to encourage users of MovieLens, a movie recommendation website, to rate more movies. They show that providing users with information on how their rating output compares to others' substantially increases the volume of ratings. They also find some evidence that a user's propensity for altruism predicts their likelihood of rating movies that have few existing ratings. This is significant for the current paper, as it suggests that at least some users are motivated by altruism when providing ratings.

It is important to distinguish the current line of research from several papers that have been written on two-sided reputation systems. Houser and Wooders (2005), for example, examine the impact of reputations in eBay auctions, in which