Dynamic trends in online product ratings: A diagnostic utility explanation☆

Fang Wang, Kalyani Menon[1], Chatura Ranaweera[*],[1]

Lazaridis School of Business, Wilfrid Laurier University, 75 University Ave West, Waterloo, ON N2L 3C5, Canada

ARTICLE INFO

ABSTRACT

We propose and test an alternative diagnostic utility explanation for trends in customer ratings of products on online reviews/ratings sites. Ratings by prior customers provide key diagnostic information to help prospective customers in their purchase decision. The outcomes of these purchase decisions are reflected in the ratings posted by these later customers. Therefore, rating trends depend on the diagnostic utility of the ratings environment. We suggest that this diagnostic utility is a function of the degree of heterogeneity in the ratings environment and customer diagnostic ability. Using two data sets - from Landsend.com and from Amazon.com, we show a predominant increasing rating trend and support for our diagnostic utility explanation. Ratings show an increasing trend unless heterogeneity in the ratings environment increases and customer diagnostic ability decreases.

1. Introduction

Not only are customer reviews/ratings on online product review sites a crucial source of information for potential customers, significant research across a range of industries also shows positive relationships between online reviews/ratings and immediate and long-term firm performance (Floyd, Freling, Alhoqail, Cho, & Freling, 2014; You, Vadakkepatt, & Joshi, 2015). An important feature of online reviews/ratings sites is that they present a visible trace of the *trends* in online ratings. Specifically, most sites provide aggregated customer ratings, the distribution of ratings, the history of ratings, and qualitative reviews, for extended periods of time (Jin, Hu, & He, 2014). These trends are important predictors of sales and stock performance (e.g., Duan, Gu, & Whinston, 2008; Fan, Che, & Chen, 2017; Godes & Mayzlin, 2004; Moon, Park, & Seog Kim, 2014; Tirunillai & Tellis, 2012). Rating trends reflect how customer experience evolves over time, influences firm reputation, and provides quick and direct market and customer feedback to marketers, big and small (DeMers, 2014; Hult, Morgeson III, Morgan, Mithas, & Fornell, 2017).

Existing research suggests that rating trends may be driven by characteristics of a website's prevailing ratings environment and/or characteristics of reviewers, and has focused exclusively on identifying and explaining declining trends (Godes & Silva, 2012; Li & Hitt, 2008; Wu & Huberman, 2008). The occurrence of and conditions for non-declining rating trends remain unexplored. This gap in the literature is of significance because non-declining rating trends signal a favorable

business environment, and understanding and enabling non-declining rating trends is therefore important.

We build on past research that views online ratings as an important information source for potential customers making purchase decisions (De Langhe, Fernbach, & Lichtenstein, 2016; Godes & Silva, 2012; Moe & Trusov, 2011). Reviews and ratings serve a diagnostic utility (Filieri, 2015; Karimi & Wang, 2017), i.e., they provide useful information to help future customers determine the suitability of a product (or service) for their own purchase needs. Greater (weaker) diagnostic utility of existing ratings should account for better (poorer) purchase decisions by subsequent customers. Better (poorer) purchase decisions by later customers will cause higher (lower) ratings by them, leading to an increasing (declining) trend in ratings. Thus, depending on the diagnostic utility, ratings can potentially show either declining or non-declining trends. This raises the question – what determines the diagnostic utility of online ratings for potential customers, and the consequent rating trends?

Past research shows that the diagnostic utility of a review site, and therefore the rating trends, may be influenced by characteristics of the review environment and/or characteristics of the reviewers. We add to this diagnostic utility theory by developing and testing a theoretical framework focusing on specific, and thus far uninvestigated, aspects of the review environment and the reviewer base - the heterogeneity of ratings in the environment and customer diagnostic ability respectively. High or increasing heterogeneity refers to low consensus in the review environment and is likely to compromise the diagnostic utility of the

---

environment. Customer diagnostic ability refers to their ability to access and interpret the posted ratings in light of their own purchase needs. Higher diagnostic ability can lead to better purchase decisions. Changes in the heterogeneity of the ratings environment and incoming customers' diagnostic abilities will affect the diagnostic utility of the existing reviews/ratings to incoming customers, leading to various trends in ratings.

We test our theoretical framework using data from Landsend.com, and conduct a robustness check using data from Amazon.com. The results confirm that customer diagnostic ability and heterogeneity in the ratings environment are important factors affecting rating trends. An increasing trend exists when diagnostic utility is not compromised – specifically, when heterogeneity and customer diagnostic ability do not change significantly over time; while such an increasing rating trend disappears when heterogeneity increases and customer diagnostic ability decreases over time.

Broadly, we contribute to the literature on customer use of online ratings in their pre-purchase information search (De Langhe et al., 2016; Punj, 2012; Wang, Liu, & Fang, 2015). More specifically, we contribute to an understanding of what drives rating trends – an important and unique but under-researched feature of online rating sites. We re-formulate the diagnostic utility explanation for rating trends by introducing two key conditions - customer diagnostic ability and heterogeneity in the ratings environment, and we consider the simultaneous effect of these two conditions. By doing so, we theoretically extend the previous explanations that, while studying diagnostic utility as the underlying force driving rating trends, focused only on declining trends. Our extension enables us to predict and explain various types of rating trends, declining and non-declining. This study is the first to document a predominant increasing rating trend as well as non-increasing rating trends, and explain both in terms of the diagnostic utility of the reviews/ratings environment.

Finally, we provide insights for the management of online ratings. The extant literature appears to question the wisdom of maintaining online customer ratings. If indeed customer ratings show only predominant declining trends, as past research suggests, then it might be counter-productive for marketers to allow customer feedback beyond a certain point. However, given the positive effects of customer-to-customer information sharing (Gruen, Osmonbekov, & Czaplewski, 2006), curtailing customer feedback can also be counter-productive. It is in this context that we provide evidence that an upward trend in ratings is in fact feasible, and provide actionable recommendations for the design and maintenance of such favorable review environments.

2. Literature review

Appendix A summarizes past literature directly examining the dynamic trend of online ratings (i.e., do online ratings systematically increase or decrease in valence over time), and positions our work in relation to this previous body of research. Our review of the literature (Table 1) reveals that trends in ratings have been attributed either to customer characteristics such as reviewer composition change (i.e., changes in who is reviewing) or to changes in the review environment

Table 1
Key previous explanations for trends in online ratings.

Causes of ratings trends	
Change in reviewer characteristics	Change in existing review environment
Customer segment changes over time. (Self-selection explanation)	The diagnostic value of existing reviews changes over time. (Diagnosticity assessment explanation)
Disagreeing reviews are over-represented in the review sequence. (Motivation explanation)	

itself.

Change in reviewer characteristics refers to the idea that distinct types of customers experience the product at distinct stages of the product lifecycle. Conceptually derived from the new product adoption curve (Rogers, 2003), distinct customer segments enter the market at different stages of the product life cycle, consume and evaluate the product, and subsequently post a rating that reflects their 'true' evaluation. The differences between the distinct segments, for e.g., the early and late adopters, account for the rating trends. This explanation is referred to as *self-selection* because customers self-select into the market for the product. For instance, Li and Hitt (2008) found that ratings on Amazon.com for books by popular authors are more likely to decline over time than are ratings for books by less productive/popular authors. Using the self-selection explanation, they suggest that popular authors have groups of avid fans who tend to purchase and review positively immediately after release, causing the early spike and subsequent fall in ratings as less avid fans enter the market.

The other potential cause of rating trends is changes in the review environment itself, i.e., change to the existing body of reviews and ratings. The *diagnosticity assessment explanation* (Godes & Silva, 2012) suggests that the increasing volume of ratings in the review environment makes it difficult for new customers to accurately use the ratings information to make appropriate purchase decisions, thus generating declining rating trends. In supporting the diagnosticity assessment explanation, Godes and Silva (2012) report that the declining rating trend is more pronounced when reviewers are highly heterogeneous. The diagnosticity assessment explanation recognizes the diagnostic utility of the ratings environment as the main driving force underlying rating trends, and is therefore of special interest to us. Building on it, in the next section we suggest an alternative diagnostic utility explanation that enables explanations for not only declining but also non-declining trends.

Finally, Wu and Huberman (2008) propose a *motivation explanation* for ratings trends where they consider both changes in the review environment as well as changes in the reviewer composition. This explanation was later tested by Godes and Silva (2012). According to this explanation, consensus (or lack thereof) in the existing ratings environment can motivate certain types of customers to post. Specifically, some customers are motivated to post a rating only if their opinion diverges from existing opinions. In other words, customers will undertake the cost of posting only if they can add value by being different. Therefore changes in the review environment and associated changes in who decides to post a review (reviewer characteristic change) determine the trends. Disagreeing opinions are increasingly over-represented in the ratings due to reviewers' motivation to post such ratings, impacting rating trends.

Common to this past literature is the focus on declining trends and post-hoc theoretical explanations for these trends. Some of these explanations seem to imply that a declining trend, if not inevitable, is definitely the dominant trend to be expected. This is attributed either to the extra marketing push or the presence of avid fans at the launch of a product that sets ratings up for a decline (Li & Hitt, 2008), or due to the heterogeneity and cognitive overload unavoidably linked to an increasing volume of ratings (Godes & Silva, 2012). The former explanation is limited to certain types of products (e.g., products with avid fans), while the latter explanation may be questioned because of literature indicating that an increasing volume of information need not in itself lead to cognitive overload and sub-optimal purchase decisions (Filieri, 2015; Jacoby, Speller, & Berning, 1974; Jacoby, Speller, & Kohn, 1974; Lurie, 2004). Empirically this stream of research may have been subject to a ceiling effect in that the declining trend reported by this past research is, among other things, unduly affected by the specific sampling environments. These studies feature samples of relatively high average ratings, ranging from 4.02 to 4.14 out of 5. The high average ratings and consequent ceiling may indicate that ratings are more inclined to go down, whereas a less skewed data environment will be less