Navigating By the Stars: What Do Online User Ratings Reveal About Product Quality?

Bart de Langhe, University of Colorado at Boulder
Donald Lichtenstein, University of Colorado at Boulder
Philip Fernbach, University of Colorado at Boulder

We show that user ratings are only weakly related to product quality and far less diagnostic than price. Yet, consumers mostly rely on user ratings and much less on price. Consumer trust in user ratings is largely misplaced because these ratings are biased by brand reputation and price.

[to cite]:


[url]:

http://www.acrwebsite.org/volumes/1017242/volumes/v42/NA-42

[copyright notice]:

This work is copyrighted by The Association for Consumer Research. For permission to copy or use this work in whole or in part, please contact the Copyright Clearance Center at http://www.copyright.com/.
Navigating by the Stars: What Do Online User Ratings Reveal About Product Quality?

Bart de Langhe, University of Colorado, USA
Philip M. Fernbach, University of Colorado, USA
Donald R. Lichtenstein, University of Colorado, USA

EXTENDED ABSTRACT

Consumers typically have to make an inference about a product’s quality before buying. Traditionally consumers have drawn on marketer-controlled variables like price and brand name to make such inferences (Rao and Monroe 1989), but the consumer decision context has changed radically over the last several years. One of the most important changes has been the proliferation of user-generated content (Keen 2008). Almost all retailers now provide user reviews and ratings on their websites, and these reviews often play an important role in the purchase process (Mayzlin, Dover, and Chevalier 2012). The steep rise of user ratings suggests that consumers think they are valuable for making good decisions. But just how much faith should consumers put in these ratings?

We identify a number of potential challenges to the predictive validity of user ratings for quality. Previous research indicates that ratings do not faithfully represent the population of users (Mayzlin, Dover, and Chevalier 2012, Hu, Pavlou, and Zhang 2006), and raters are influenced by previously posted ratings, creating herding effects (Moe and Trusov 2011; Muchnik, Aral, and Taylor 2013; Schlosser 2005). Moreover user experience is highly variable and there is often not a sufficient sample size to yield a precise estimate of the average. Another concern is that product quality is often hard to discern from user experience and evaluations are therefore biased by marketing variables like price and brand name. This raises the possibility that user reviews are similarly biased.

We developed a database of (1) quality scores from Consumer Reports, (2) user ratings from Amazon.com, (3) selling prices, and (4) brand perceptions from a proprietary industry survey. Analyses reveal that the average user rating is about equally predictive of quality as the number of ratings and much less predictive than price. Not controlling for other predictors, having a higher Amazon user rating predicts having a higher quality score only 57% of the time. Differences in user ratings lower than 0.40 have essentially no predictive value and larger differences predict quality superiority only about 65% of the time. We also found that the predictive validity of the average user ratings depends substantially on its precision as measured by standard error. As sample size increases and/or the variability of the distribution decreases, user ratings become better predictors of quality. Unfortunately a large proportion of products in the marketplace do not have sufficient precision to yield a reliable indicator of product quality. We also examined the predictive validity of price and found it to be similar to what has been suggested in past research (Tellis and Wernerfelt 1987). Notably, although the correlation between price and quality is modest, it is a much better predictor than either the average or number of ratings.

In a second study we asked consumers to search for pairs of products on Amazon.com and then to rate the relative quality of these products. Consumers perceived a significant positive relation between all three cues—price, average user rating and number of user ratings—to quality. Average user rating was by far the strongest predictor of quality perceptions. We also found that consumers fail to moderate their reliance on the average user rating as a function of its precision. Although the effect of average ratings on quality inferences may not be surprising in light of many studies documenting their pervasive influence on consumer and managerial decision-making, it is notable that consumers think the average user rating is so much more predictive of quality than price. The price-quality heuristic is one of the most studied phenomena in the marketing literature and typically considered to be one of the cues that consumers treat as most diagnostic of quality. Our study is the first to revisit the strength of the perceived price quality relationship in a multi-cue environment where user ratings are simultaneously present, which is a common mode in which consumers evaluate products today.

Study 1 revealed that average user ratings are not very diagnostic of technical quality. We hypothesized that one reason for this is the biasing influence of marketing actions on user evaluations. To test this idea in Study 2 we supplemented the market data with brand perception data and conduct a new regression that predicts average user rating as a function of price, brand perceptions and technical quality. By controlling for quality we isolate the contaminating influence of extrinsic cues on user ratings. As predicted, the dissociation between user ratings and quality can be traced in part to the influence of marketing actions on users’ evaluations. Controlling for quality, average user ratings are higher for more expensive products, higher for brands with a reputation for offering more functional and emotional benefits, and lower for brands with a reputation for affordability.

We conclude that consumer trust in online user ratings is largely misplaced. It may be a kind of “illusion of validity” where the predictive accuracy of a cue is overestimated due to its resemblance to the outcome variable, regardless of factors that undermine its value (Tversky & Kahneman 1974).

REFERENCES


Keen, Andrew (2008), _The Cult of the Amateur: How blogs, MySpace, YouTube, and the rest of today’s user-generated media are destroying our economy, our culture, and our values_. Random House Digital, Inc.


