DO I FOLLOW MY FRIENDS OR THE CROWD? INFORMATION CASCADES IN ONLINE MOVIE RATING

Young Jin Lee • Yong Tan • Kartik Hosanagar

Austin E. Cofrin School of Business, University of Wisconsin, Green Bay, WI 54311
Michael G. Foster School of Business, University of Washington, Seattle, WA 98195-3226
The Wharton School, University of Pennsylvania, Philadelphia, PA 19104
leeyj@uwgb.edu • ytan@uw.edu • kartikh@wharton.upenn.edu

Abstract

Online product review as a form of online Word of Mouth (WOM) and User-Generated Content (UGC) has attracted much attention recently. While there are many studies relating online reviews and product sales, the interesting and important problems regarding user review generation processes have been largely ignored. This study analyzes how online movie user ratings are generated through a complex interrelationship between product information, marketing effort, and social influences. In particular, we examine the effects of comparable observational learning from the crowd or friends on user ratings. This study exploits sequential user movie ratings in an online community with user and movie level information, and constructs plausible latent variables for users' perceived movie quality and the heterogeneity at movie and user levels. Our analysis indicates that, on average, higher predecessors' ratings increase the likelihood of a subsequent user providing a high rating; in other words, herding occurs. On the other hand, the degree of herding behavior by prior friend ratings becomes relatively smaller. More interestingly, the impact of predecessors' ratings becomes weaker as the volume of friend ratings increases. This study contributes to the understanding of how social imitation and learning affect user rating generation and how online social interactions moderate inefficiency in product quality information created by online users.

Keywords: Online Word of Mouth, Informational Cascades, Observational Learning, Herding, Latent Variables, Multilevel Model, Movies

1. Introduction

Online user generated reviews, as a reliable source for quality information about experiential products, can significantly enhance the "Buzz" effect surrounding the products. In recent years, firms have deployed new services and business models to leverage user-generated reviews. In March 2009, Netflix integrated its web application with Facebook to let users link their accounts at the two sites and share movie user ratings (Tirrell 2009). Under this scheme, movie ratings on Facebook will be linked back to the corresponding movie pages on Netflix, so other members can easily discover movies and add them to their own queues. Similarly, some startups have been developing new recommendation services by aggregating scattered user generated reviews across various online communities. For example, Nanocrowd analyzes user generated reviews and ratings data from various sources on the Internet and creates a list of customized movie recommendations for users (Glockner 2009).

Existing work in this area has analyzed the design and performance of eBay and Amazon-like online recommendation systems (for example, see a survey in Dellarocas 2003). Specifically, most studies on online reviews focus on the ex-post valence and dispersion of online reviews and their relationship with sales. There are mixed findings in the literature on how online user reviews of a product influence its subsequent sales (Godes and Mayzlin 2004, Chevalier and Mayzlin 2006, Duan et al. 2009, Dellarocas et al. 2007). The positive impact of user reviews on book sales for online retailers such as Amazon.com or BarnesandNoble.com has been empirically tested (Chevalier and Mayzlin 2006, Li and Hitt 2008). Chen et al. (2004), however, reports no impact of user ratings on sales from a similar data source of Amazon.com. Duan et al. (2005) and Liu (2006) examine the positive impact of online review volume on sales. Recently, Li and Hitt (2008) analyze how idiosyncratic preferences of early buyers can affect long-term consumer purchase behavior through online review systems, and show that altering marketing strategy to encourage consumers to generate positive reviews to self-select into the market early could be beneficial for new products. Forman et al. (2008) find the stronger association between reviews and sales in the presence of reviewer identity disclosure. Under informational cascades, Duan et al. (2009) look specifically at the impact of user ratings on software adoption and find that user ratings have no significant impact on the adoption of popular software.

Another research stream is related to the motivation for users to generate reviews on the Internet. Social psychologists have for long been studying emotional sharing. People share their opinions shortly after an event occurs, with over 50% even on the same day. The majority tends to share recurrently with multiple people. There are a few differences in the extent of sharing with regard to age, gender and culture (for reviews, see Rimé et al. 1991, 1992, 1998). Peer recognition is a positive motivator since people value such recognition (Jeppesen and Fredericksen 2006). The work on online reputation systems has primarily focused on the consequence of peer recognition (Resnick et al. 2000). Self-verification is

another important driver of online contribution. People wish to feel connected to others in social groups and receive identity-affirming feedback from others in the group (Jones and Pittman 1982, Swann 1983, Brewer 1991, Baumeister and Leary 1995, Tyler et al. 1996, Hornsey and Jetten 2004). Regardless of the motivations of user generated reviews and their relationship with product sales, intuitively, positive online review posts encourage other consumers to adopt products whereas negative opinions discourage them (Dellarocas et al. 2007).

These studies do not, however, directly analyze how online user reviews are generated and whether they effectively convey and move information about quality through large-scale consumer networks. In addition to directly observable product information and surrounding environmental characteristics such as competition and advertising, we consider two major influential factors to examine online user review generation and quality information flow. First, the opinions created by preceding users (other reviews) may affect the current user's perception on quality. Especially, the components of social influence such as word of mouth, product information diffusion, network effects, and informational cascades, are intermingled and it is difficult to clearly identify each component (Duan at al. 2009). People frequently engage in "observational learning," drawing quality inferences from observing peer choices (Zhang 2010). It is possible that the observation of choices often coexists with other sources of quality information such as word of mouth communication (Ellison and Fudenberg 1995, Godes and Mayzlin 2004, Mayzlin 2006), network externalities (Nelson 1970, Erdem and Keane 1996, Camerer and Ho 1999, Villas-Boas 2004, Hitsch 2006, Narayanan et al. 2007), and firms' marketing mix variables (Moorthy and Srinivasan 1995, Wernerfelt 1995, Desai 2000, Anderson and Simester 2001, Guo and Zhao 2009, Zhang 2010). People may overreact towards more positive (negative) reviews or be disgruntled by other reviews, but the communication or discussion between friends (via social networks, e.g., online friends are acquainted with each other in an online or offline community) may lead to private information flow. If a consumer can clearly identify her friend as a very conservative person in rating after communication, this friend's rating could provide different quality representation. Also, people may or may not be attracted by intensive advertising and it is unclear how online user ratings are related to the advertising effort by firms. As such, it remains unknown how and to what extent prior reviews transfer information and private opinions about the product to subsequent adopters (or reviewers) under a variety of observable product quality measures. This becomes more complicated when each review is summarized into a numerical value and its underlying tone or reason cannot be easily revealed. In addition, manipulation of ratings by firms and reviewer self-selection bias can distort information from observable online user ratings (Li and Hitt 2008). Therefore, online reviewers encounter a problem of rating decision under imperfect information from prior user reviews and their own private signals.

Second, users' different tastes or preferences can be fundamental identifiers when online reviews are generated. The complexity and difficulties in this research is mainly attributed to the unobservable consumer preference of product and motivation of reviews. Different preferences will provide different thresholds to consumers in product adoption and even evaluation decisions. This problem becomes more sophisticated when the thresholds are affected by others' decisions or similarity in social systems. For example, individuals in the same reference group may tend to behave similarly in a common environment. Hence, it is difficult to distinguish real social effects from correlated effects, known as the reflection problem (Manski 1993, Bramoullé et al. 2009). In order to correctly examine the generation of online user reviews, another challenge is laid on isolating social influence from consumer heterogeneity, homophily (McPherson et al. 2001), and other relevant information such as product characteristics and firms' marketing mix variables.

In this study, we characterize user product evaluation behavior in online movie review generation and examine observational learning in sequentially generated ratings. In particular, we use publicly available data from an online movie social networking community website—Flixter.com (Flixster). Other data such as movie characteristics, performance, and marketing spending are collected from various sources to meet the challenges described above. Flixster is the largest and fastest-growing online movie community. It has over 20 million unique users who use its website and Facebook application each month and about 2 billion movie reviews created by users. Flixster provides its web and mobile applications across a wide network of popular online platforms like Facebook, MySpace, Bebo and iPhone. This facilitates users to share their thoughts and opinions with friends instantly as well (Dobuzinskis 2009). It is ideal to examine observational learning in this online movie community for the following reasons. First, user ratings are time-stamped, and the sequence of user ratings is easily identified. Second, a unique user name associated with each rating and the user's friendship in the community clearly separate observable ratings by the crowd and friends. The enhanced features of Flixster's web or mobile applications not only increase the visibility of friends' ratings and sharing opinions but also make it very easy to read the corresponding text reviews (See Figure 1). Meanwhile, the likelihood of within-crowd communication is minimized but the possibility to communicate between online friends is maximized. Hence, the identified ratings in friendship network among users may mimic true online WOM and distinguish observational learning without and with online communication, which in turn enables us to investigate comparable information-motivated herding (Banerjee 1992, Bikhchandani et al. 1992) in online movie ratings. Hence, online user reviews and ratings are quality information indicated by prior reviewers, which may drive later reviewers to generate similar ratings (Li and Hitt 2008). Third, similar to Zhang (2010)'s evidence of observational learning in patients' sequential decisions of kidney transplant, online users are unlikely influenced by other primary mechanisms behind uniform social behavior, such as sanctions of deviants,

preferences for social identification (Kuksov 2007), and network effects (Yang and Allenby 2003, Nair et al. 2004, Sun et al. 2004). In particular, online users' rating choices do not usually contain public appearance value that professorial critics mostly concern about. Online reviewers do not usually derive additional benefit from generating reviews and, hence, there is no apparent positive externality existing in product adoption (Katz and Shapiro 1985). The U.S kidney market in Zhang (2010) presents that the awareness of transplant kidneys does not play a significant role due to the limited number of donors and patients. However, movie information diffusion may be an alternative factor that possibly leads to herding behavior (Duan et al. 2009) in online rating participation and choices. For example, the likelihood of sharing opinions and participating in online reviews increase with the number of previous online reviews; this is similar to the traditional product information diffusion (Bass 1969), regardless of the overall valence of ratings.

| A. Movie and User's profile page | B. Rating Page |
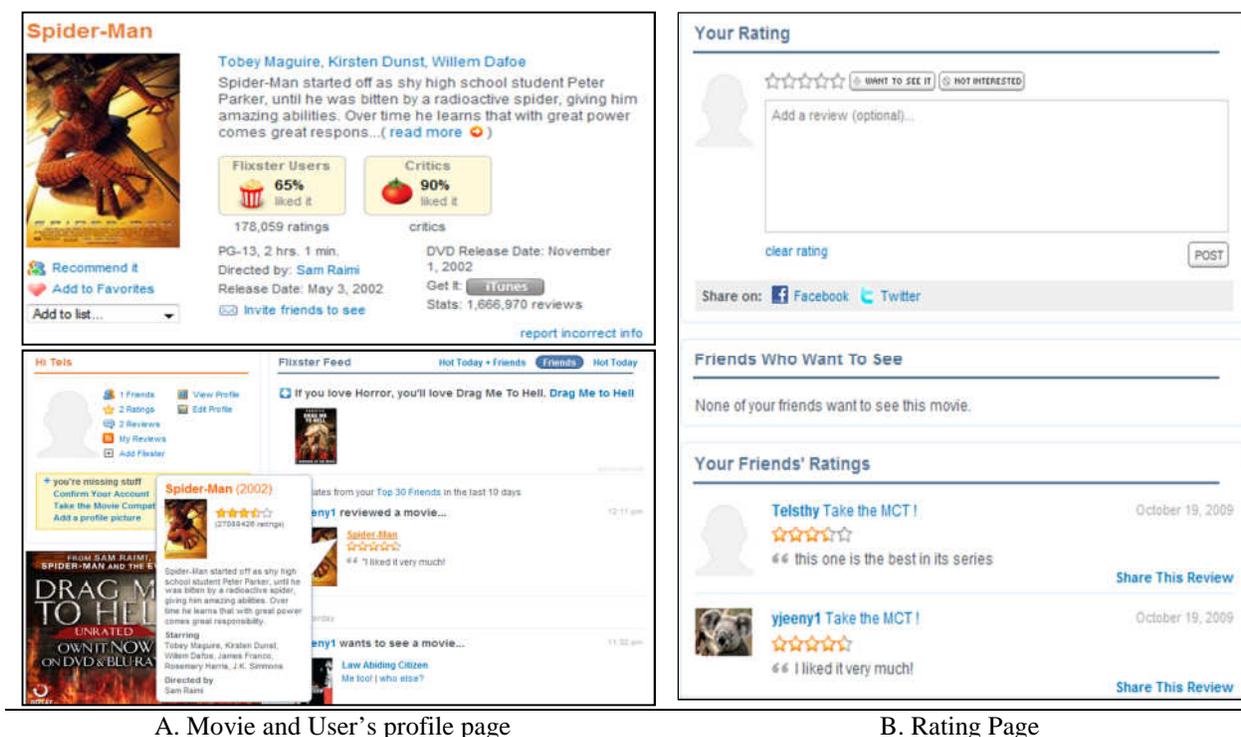
Figure 1. Screen shots of a user profile and movie webpage in Flixster.com

Our analysis is designed to examine observational learning and herding behavior in online movie user ratings, as suggested by the informational cascades theory (Banerjee 1992, Bikhchandani et al. 1992, 1998). While all moviegoers can observe public movie related information (e.g., genre, MPAA rating, box-office rank and sales, critics' reviews, advertising, and etc.), each audience creates perceived quality about a movie (e.g., private quality signals (Zhang 2010)) after watching it. If an online user wants to share her opinion about a movie at an online community, she may observe others' prior opinions before or

after watching the movie (e.g., ratings by other prior users or friends) in the community. Although each prior user's rating is virtually observable, the vast amount of reviews available on the community has created information overload among online users (Shapiro and Varian 1999, Brynjolfsson and Smith 2000, Jones et al. 2004). An online user may infer the quality of a movie only collectively by the average prior rating for the movie (Li and Hitt 2008). In particular, while the average rating is a measure without any interaction with other users (i.e., observational learning from others in rating is enabled as illustrated in Figure 1-A), the quality signal from prior ratings by her friends may present a different characteristic, having both the friends' ratings and private reasoning (e.g., text reviews corresponding to numerical ratings or online chatting with the friends, See Figure 1-B), therefore the friends' ratings and text reviews are more similar with offline WOMs. As a result, a user's rating decision depends possibly on others' rating decisions as well as her private quality information.

In the observational learning literature, it is proved theoretically that an individual can rationally follow the behavior of preceding individuals without regard to her own information, having observed the actions of those ahead of her—*informational cascades* (Bikhchandani et al. 1992, p. 994). *Herding* describes a phenomenon in which individuals converge to a uniform social behavior (Banerjee 1992, Bikhchandani et al. 1998). Several empirically oriented studies (e.g., Anderson and Holt 1997, Çelen and Kariv 2004a) demonstrate the convergence in individuals' actions in experimental environments. In this paper, the time series of user ratings help to identify observational learning in online user rating from the different sources (the crowd or friends) of prior ratings since the time position of a user's rating indicates the amount of observational learning the user is exposed to (Zhang 2010). As a result, instead of testing user rating convergence, we are able to examine the impact of observational learning on user rating by applying latent response approach at the individual level. In particular, we explore the presence and implications of comparable observational learning in online user rating by addressing the following research questions:

- Does observational learning exist in online user rating even when a user has already created private quality information?
- Is a user influenced by others' ratings differently based on the type of observational source when making a rating choice?
- How does the observational learning relate to social interactions among users?
- How do online user characteristics (e.g., age, centrality in a social network, the amount of online activity, visibility to others, etc.) affect rating behavior?
- How should firms and recommendation system designers adjust their strategies to account for possible user behavior in online rating?

The extant studies (Chevalier and Mayzlin 2006, Li and Hitt 2008) have found the evidence of positive impact of online user rating on online book sales while the impact is insignificant in popular software adoption (Dual et al. 2009). To gauge the strength of our results in observational learning in user rating generation, it is critical to show that online user movie rating is positively linked to subsequent box-office sales. Hence, our sales model is exclusively designed to test this economic evidence after we analyze observational learning. Actual box-office sales of population-level movies in 2007 and their advertising expenditures are used to estimate our sales model. This differentiates our study from the extant studies that have used a subset of products or sales channels, and proxies for promotions (Chevalier and Mayzlin 2006, Li and Hitt 2008). In particular, we examine the impact of online user rating on weekend box-office sales by addressing the following two questions:

- If we are able to separate average online user rating for a movie into a component from all other factors, how does it affect the long-term weekend box office sales of the movie?

- How is the impact of online user rating different between all movies and relatively popular ones?

By applying several estimation methods to address heterogeneity in user and movie levels and possible homophily, our findings suggest that, on average, higher predecessors' ratings increase the likelihood of a subsequent user to provide a higher rating; in other words, herding occurs. We also show that the degree of herding can be significantly different across movies due to movie level heterogeneity. On the other hand, the degree of herding due to prior friend ratings becomes relatively smaller. More interestingly, the impact of predecessor's ratings becomes weaker as the volume of friend ratings increases. Hence, this study contributes to the understanding of social imitation and learning in online user product recommendation and how online social interactions moderate inefficiency in product quality information created by online users. Finally, in a separate analysis, we find strong evidence of positive impact of average user rating on the long-term box-office sales and compare it to the effect of advertising to highlight the significance of online user rating and the implication of observational learning in the user rating generation process. Therefore, our findings provide useful managerial implications for social media marketing and recommendation-system design as well.

We organize the rest of the paper as follows: First, we introduce research background and formulate the hypotheses. Second, we describe the data. Third, we examine user rating generation behavior and social influences by several estimation techniques. Fourth, the empirical evidence of positive relationship between online user rating and long-term box-office sales is presented with robust econometric methodology. Finally, we discuss the implications and offer directions for future research.

2. Research Context and Hypotheses Development

Our study is established on the user rating generating process of the sequential choices of users who interact with other users primarily in an online movie community, Flixster. Flixster presents an ideal

environment for this study. Virtually, all movies' (e.g., upcoming and current in theaters and DVDs) basic information is available at Flixster by simple search. On each movie page, users are presented with a wide array of movie information: synopsis, actors, directors, running times, release date, MPAA rating, and etc. In addition, several types of movie recommendation are available on each movie page at Flixster: information on average user rating, average critic rating, and each friend's rating with an attached text review, and total number of reviews for a movie (See Figure 1). A user can recommend any movie link directly to her friends and forward a friend's review easily to other friends in Flixster and even via other Web applications (e.g., Facebook and Twitter). For information on user ratings and reviews, Flixster updates the cumulative number of ratings and friend rating associations with individual users (e.g., friend ratings will be presented separately and immediately) for each movie every day. Like other popular social networking websites (e.g., Facebook and MySpace), each user can create her own profile webpage in Flixster. A user can manage her online friendship and other users can visit her online profile webpage (e.g., name, gender, location, membership duration, the number of views by others, total number of friends, the number of generated ratings, the number of generated text reviews, and etc. See Figure 1-A).

In addition, public movie quality information should be controlled for. First, the time series of movie performance information reflects the total number of adopters (audience in theaters) and the extent of product diffusion (Bass 1969) at a given time. Box-office sales and ranking of each movie at a given time are publicly available at movie information websites such as BoxOfficeMojo.com and Numbers.com. Second, another piece of public movie quality information is critics' rating. Unlike sales and ranking, movie critics' ratings (e.g., Metacritic.com) are usually presented at the beginning period of a movie's release, therefore average critic's rating is time-invariant in our study. Third, potential audience (even after watching) can be exposed to subjective quality information from firms. The advertising effort is intense right before the movie release day. However, firms can allocate marketing budget even after movie release (Dellarocas et al. 2007). This alternative quality inference from advertising should be controlled for as well. Hence, we have collected weekly advertising spending for each movie from TNS Media Intelligence. Our rich data sources combine all possibly observable explanations of user quality inference and distinguish our study in identifying observational learning from the extant informational cascades literature.

2.1 Informational Cascades in User Ratings

Generally, a user watches a movie in theater if her expected quality of the movie is greater than its ticket price. Then, the users may want to express and share opinions online about the movie. That is, a user's movie rating choice involves several phases of quality inferences (See Figure 2). For example, based on all prior quality inferences before watching a movie (by ads, own valuation, peer and expert reviews), only users who have expected utility greater than about $10 would watch the movie. Though the expected

quality of a user would be updated into her perceived quality after watching a movie, she still faces another decision to choose a rating value (e.g., one of numerical values or stars in a 1 to 5 or 10 scales). As a result, the user has two sources of information in generating a review online for the movie. One is her own information based on the perceived quality of the movie. Although the information can be more certain than the expected quality before watching the movie, she can still be uncertain about the true rating value. The other is the information derived from rating choices by others (Duan et al. 2009) – two comparable ratings from the crowd and friends in our study. Others' prior ratings can still possibly affect not only a user's decision to watch a movie (Li and Hitt 2008) but also her rating decision. She then combines the two sources of information (e.g., self-perceived quality and prior other ratings) to make her rating choice about the movie. Consequently, we propose one following hypothesis to evaluate our assumption that a subsequent user is influenced by other users' ratings when she generates a rating even after watching a movie.

HYPOTHESIS 1 (H1): Herding Behavior in User Rating. *The higher the average rating by prior users is, the more likely a subsequent user is to choose a high rating versus a low rating.*
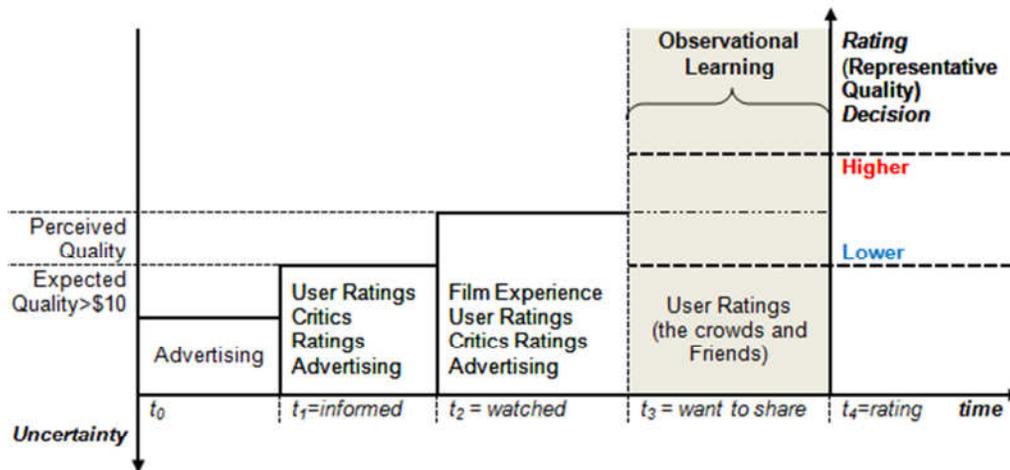
Figure 2. A time line of a user *j* to rate a movie *i*

Underlying observational learning from the two comparable ratings of the crowd and friends may send different quality signals although the observations (e.g., rating valences) are the same. While both ratings are regarded as the actions (or choices) of others which may trigger observational learning, they represent quality differently in the context of WOM to impact subsequent users' actions (Katz and Lazarsfeld 1955, Coleman et al. 1966, Gladwell 2000). Online community (Sproull 2003) resembles offline community not just in the way of communication activities (e.g., email, blog, and instant messaging) but also social activities (e.g., making friends, establishing self-recognition or social status in social networking sites). Hence, WOM flows in online community similarly as in offline world. Also,

online WOM is not created or consumed equally and its impact depends on who is talking to whom (Godes and Mayzlin 2004). Unlike ratings from the crowd, the observed ratings from friends can have embedded WOM effect due to online communication and interactions among people which is fundamental in the notion of WOM (Arndt 1967). WOM effect is a different social force to impact consumer action compared to informational cascading in which users rationally ignore, or weigh less on, their private information (any quality inference) and generate similar actions of others (Bikhchandani et al. 1992; Vives 2008). Hence, the choice behavior in rating can be influenced by a reviewer's own taste and the tastes of others (Yang and Allenby 2003).

Our study posits to identify the difference, if any, between the two comparable types of observational learning – the collective information about other users' ratings and the information about friends' ratings which embed WOM effect and bounded memory. The latter is considered "smooth" nature of learning rather than "discrete" nature which leads eventually to herding behavior (Vives 2008). When a user comes to the website, she usually can see the average rating by previous users about a movie without detailed information about how each predecessor has rated the movie. In contrast, the underlying reasons of friends' ratings are more visible by associated text reviews and other means of communication since the design of social networking sites typically makes it easier for a user to see what her friends do and change within the site. When a user rates a movie, her decision would be that of choosing one rating category over other rating categories. In this setup, rather than testing rating convergence towards the end of queues, it is important to see the evidence of observational learning from other ratings with and without sharing private information among users. For example, there may be a higher probability that a user chooses a high rating than a low rating if the observed ratings of the crowd are high when other explanations of quality inference are fixed. On the other hand, based on the informational cascades theory (Banerjee 1992, Bikhchandani et al. 1992, 1998), the probability should become relatively lower with friends' ratings due to information sharing. In other words, herding behavior is moderated due to sharing the underlying reasons of friends' ratings, which may reflect the likelihood of convergence to an efficient quality information level (Zhang 2010). In our paper, we separate the average ratings of precedent users (*CROWDRA*) and precedent friends (*FRWOM*) since each would form different quality information delivery in the aggregate statistics as described earlier. We therefore propose:

HYPOTHESIS 2A (H2A): *CROWDRA, all else equal, leads to a higher degree of herding behavior in a subsequent user's rating than the average friend rating*

HYPOTHESIS 2B (H2B): *Due to the amount of information sharing among friends, a subsequent user's herding behavior by CROWDRA is reduced by the volume of friends who rated the same movie.*

There are two challenges in this study as described earlier. First, individual heterogeneity in terms of taste, preference, and perceived quality cannot be observed. Second, other possible social effect such as

homophily should be controlled for. In our context, if there were homophily effect, it would cause a reflection problem (Manski 1993, Bramoullé et al. 2009). We address these issues in the following sections.

2.2 Other Public Quality Information: Marketing Efforts and Critics' Ratings

Beside others' rating choices, public quality information about movies may have a significant role in user rating generation. Users may end up with similar rating choices due to common context factors such as firms' marketing efforts and critics' ratings. Hence, testing observational learning in user rating might be spuriously attributed to social contagions without accounting for these variables (Zhang 2010). Many studies in recommender system literatures (e.g., Ba and Pavlou 2002, Bajari and Hortacsu 2003, Chevalier and Mayzlin 2006, etc.) have found that user feedback significantly affects the future performances (price and sales) of products across several industries. However, they have overlooked how the feedback is influenced by marketing efforts, expert reviews, and interactions among users. Hence, we also include observed advertising expenditure and the average ratings of movie critics of each movie as the proxies for public quality information in our framework. Potential moviegoers may usually consume these measures for adoption decision (e.g., buying a movie ticket). After watching movies, these measures are not very new anymore to moviegoers. As a result, an audience's perceived quality can be created by mostly actual film experience, rather than advertising information which is used for expected quality (Kirmani and Wright 1989). Since advertising is not complementary to any rating information that rating users can mimic, its informational impact on a user's rating decision problem should be minimized in our study. Therefore, we propose:

HYPOTHESIS 3A (H3A): *A total advertising effort, all else equal, is not correlated with a user's rating.*

In contrast to advertising, although users consume critics' ratings for their expected movie quality similarly as advertising, critics' ratings are comparable to user ratings. That is, critics' ratings may become an easy reference for rating users. However, critics' ratings should not be included in user rating decision queues since movie experts have generated the ratings before or at the beginning period of movie release rather than across time. The low correlations in Table 2 (0.22 between average user and critic ratings; 0.063 between individual user ratings and average critic ratings) may support our argument that critics' ratings are not positively or negatively related to sequentially generated user ratings. Therefore, we propose:

HYPOTHESIS 3B (H3B): *The average rating of critics, all else equal, is not related with a user's rating.*

2.3 Impact of User Rating on Box-Office Sales

We test the evidence of relationship between average user rating and following weekend box-office sales. This may emphasize the impact of observational learning in online user rating on long-term sales. In other words, increasing the valence of user ratings in the earlier period of movie screening may be an optimal strategy for movie distributors. Therefore, we test:

HYPOTHESIS 4A (H4A): *Weekend box-office sales of a movie, all else equal, are positively related with its prior average user rating.*

HYPOTHESIS 4B (H4B): *The impact of average user rating on weekend box-office sales of more relatively popular movies is greater than the average movies.*

Since our observational learning study uses a subset of population level of movies and sales in 2007 and the subset includes relatively popular movies based on the volume of user reviews in 2007, H4B will further validate the importance of our main analysis.

3. Data Collection

Our dataset contains movie-level characteristics and performance, and individual online user review-level information. First, using software agents, we collected movie-level data from several public websites (See Table 1 for data sources) and sampled all movies released in theaters in 2007. In addition, the data includes weekly advertising spending for each movie from Ad$pender, which is a part of TNS media Intelligence/CMR. Second, we collected user-review level data from Flixster.com based on the movies (See Table 1). All observable information of each user (who has generated at least one movie rating for the movies) is downloaded from her profile page in Flixster. Flixster provides the information of friendship among users. Hence, it enables us not only to collect individual specific information such as gender, age, the number of generated ratings and reviews, and profile status (the number of profile viewed by other users and membership duration) in the website but also to partially observe friendship networks among users on the movies. We first considered the intersections of our movie-level and user-level review datasets. Then, for our main analysis, we selected a subset of the movies in which each movie have more than 1,000 numbers of user reviews within the first four months of the release in order to capture enough variations across rating users (See Table 2).

The 17 sample of the movies are relatively popular ones in 2007. Although the proportion of the sample movies is about 5% of all the movies in 2007, they correspond to about 25% of all user reviews for all the movies, while in theaters, in 2007 (See Table 3). They account for about 36% of the total weekend gross sales of all movies in our dataset. Table 3 provides a summary of user ratings for all movies and the 17 sample movies. While the average number of user rating for the sample is five times greater than that for all the movies, the average and variation of user ratings in the sample is not very different. Hence, the 17 sample movie dataset provides a good subset to examine observational learning

due to a larger number of ratings per each movie and screening week. Most reviews were generated within the first 12 weeks after the movies were released as shown in Figure 3. As a result, our final dataset for individual user level analysis contains 20,473 individual users who reviewed and rated at least one of the 17 movies (as shown in Table 4) and 30,628 rating observations.

Table 1. Online Users & Movies Released in 2007

Data Level	Dimension	Data Sources
Movie Level - 395 Movies in 2007 - 178,958 user ratings in Flixster.com during showing periods of the movies in theaters	Movie characteristics Weekly advertising Weekly performance	BoxOfficemojo.com / ImDb.com/ Numbers.com /Metacritic.com Ad$pender (TNS Media Intelligence) BoxOfficemojo.com Flixster.com (Ratings and Reviews)
Online Movie Reviewer Level - 63,764 individuals in 2007	Demographic & Online Profile Friendship Rating & Text Review	Flixster.com

Table 2. The Sample of 17 Movies and Average User and Critic Ratings

ID	Title	Users	Critics	ID	Title	Users	Critics
1	Bourne Ultimatum	8.46	8.5	10	Hairspray	8.77	8.1
2	Fantastic Four (2007)	6.95	4.5	11	I Am Legend	7.38	6.5
3	Harry Potter (2007)	7.99	7.1	12	Ocean's Thirteen	7.31	6.2
4	Knocked up	7.75	8.5	13	Pirates of the Caribbean (2007)	8.26	5.0
5	Shrek the Third	7.00	5.8	14	Simpsons Movie, The	7.95	8.0
6	1408	7.30	6.4	15	Spider-Man 3	7.15	5.9
7	300	8.54	5.1	16	Sweeney Todd (2007)	8.38	8.3
8	Disturbia	8.06	6.2	17	Transformers	8.59	6.1
9	Ghost Rider	6.98	3.5				-

Table 3. Summary Statistics for Number of Ratings and Average Ratings in Our Data Set

	Total Movie(371)[*]		Selected Movies(17) for first 16 weeks	
	Number of Ratings per movie	Average Rating per movie	Number of Ratings per movie	Average Rating per movie
Mean	431.76	7.02	2193.24	7.81
Standard Deviation	896.80	1.05	1524.10	0.63
Percentile (%)				
1	1	4.24	984	6.95
5	2	5.17	993	6.97
10	4	5.57	1014	6.99
25	15	6.50	1069	7.30
50	92	7.14	1455	7.95
75	492	7.77	3694	8.38
90	1153	8.29	4252	8.56
95	1758	8.50	5028	8.62
99	6037	9.00	5546	8.74

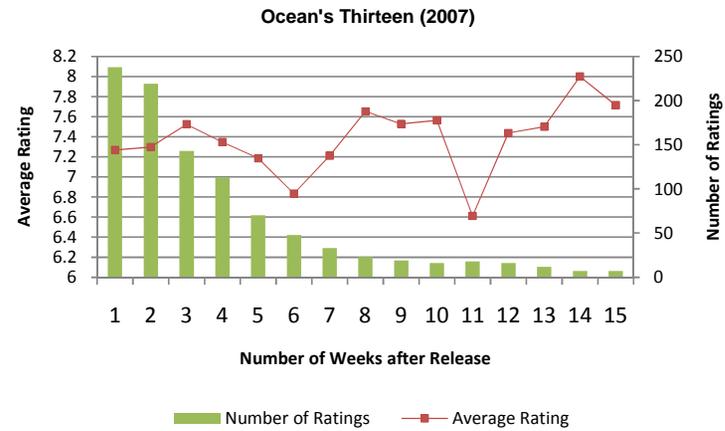[*]24 movies are removed since they have less than 5 user reviews
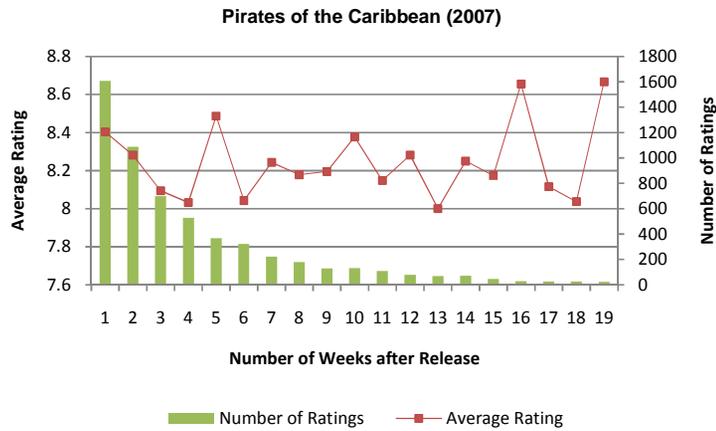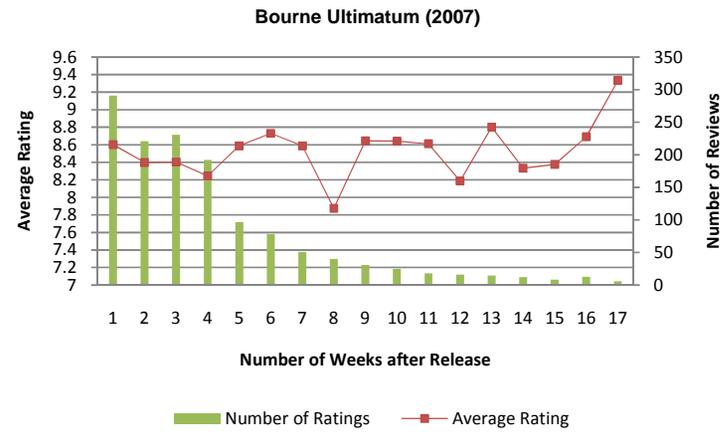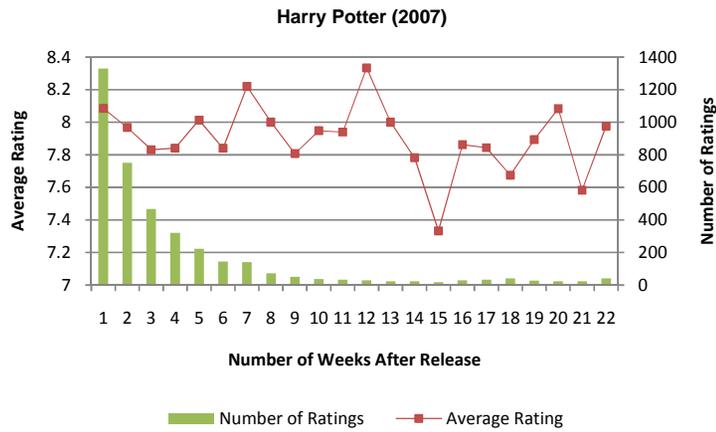
Figure 3. Online user rating trends after movies are released for 4 sample movies in our data set

Table 4. Data Descriptive Statistics

Dimension	Variable	Description	Mean	Min	Max
User Level	R	Rating for movie i	7.89	1	10
(20,473 users)	GENDER[*]	Dummy for gender (Female = 0)	0.45	0	1
	AGE[*]	Age	24.92	13	108
	MEMFOR	Membership days in Flixster	682.47	237	1286
	PRFV	Number of profile viewed by others	483.81	1	258,795
	NUMF	Number of friends in Flixster	45.02	0	830
	NUMRA	Number of ratings in Flixster	1029.43	1	68,310
	NUMRE	Number of text reviews in Flixster	164.73	1	55,667
	CROWDRA	Average prior rating by other users for movie i since it was released until time $t-1$	8.09	6.99	9.17
	FRWOM	Average prior rating by friends (6,846 obs.)	8.01	1	10
	NUMFRA	Number of prior friend ratings (6,846 obs.)	2.95	1	114
Movie Level	CR	Metacritic.com's average critic rating	6.13	3.5	8.5
(17 movies)	NR	Number of movie reviews posted on Flixster.com for movie i since it was released at time $t-1$	377.06	6	1312
	ADSPEND	Advertising spending for movie i until time t in $ million	12.70	4.03	22.73
	TOTBOXSALES	Total Gross of Box Office Sales at time t in $ million	179.4	9.30	336.53
	RD	MPAA Rating Dummy (Rated -R)	0.11	0	1
	PGD	MPAA Rating Dummy (Rated –PG)	0.72	0	1
	PG-13D	MPAA Rating Dummy(Rated-PG-13)	0.16	0	1
	WEEKS	Movie Playing Week in Theaters for movie i since it was released	4.39	1	16

[*]Since there is 8% of missing gender values in our sample individuals, we exclude the individuals of missing gender. However, 25% of included individuals have still the missing values of age. Therefore we imputed missing age values by organizing the cases by patterns of missing data so that the missing-value regressions can be conducted on other individual level variables.

3.1 Dependent Variable

The key dependent variable in our individual level analysis is an online user's choice of movie rating. Besides text reviews, users usually choose one of predefined numerical rating values to report satisfaction, recommendation, or feeling about a movie. However, the fixed scale nature of user rating does not fully represent the true degree of consumer self-reporting feedback due to discretization. Therefore, a user's internal thresholds of rating categories act as cutoff values in her rating for continuous true evaluation for a movie. On our data, each user rating contains a time-stamp of when the rating is created. This gives a great advantage to keep track of each user's rating sequential order across time. While the cumulative average user rating trends of our sample movies are very similar to the trends on online book user ratings in Amazon.com in Li and Hitt (2008)'s study (e.g., visually discernable patterns of declining and rising across times), weekly average user rating (non-cumulative) of each movie in our sample presents an up and down pattern across weeks, as shown in Figure 3. It is not clear whether the patterns are totally random or show any positive or negative relationships dynamically. Clearly, the patterns do not provide any rating convergence over time. However, we can examine statistically the evidence of observational

learning in user rating with respect to herding behavior. The average rating for the sample (17 movies) is 7.81 out of 10, which is similar to the population means reported by Chevalier and Mayzlin (2006) and the sample means by Li and Hitt (2008) which utilize aggregate level data on books.

3.2 Independent Variables: Quality Measures and User Characteristics

Our independent variables of the most interest include observable quality measures that a user can infer and individual user characteristics. The valence and volume of ratings by preceding users and friends are the most interesting variables to identify observational learning. *CROWDRA* is the average rating of all others (including online friends) who have rated the same movie. Similarly, *FRWOM* is the average rating by a corresponding user's online friends in Flixster. Therefore, zero *FRWOM* or *NUMFRA* (the number of friend ratings) indicates that a user does not have her friend rating for the same movie that she would rate. While *CROWDRA* presents the aggregate level of other user ratings without underlying information, *FRWOM* gives a user the hybrid of friend rating valence and private information (WOM) as we described earlier. Since only 22% of user rating observations have *FRWOM*, *NUMFRA* is used as an indicator of existence of friend rating when we test our hypotheses by using all observations. Hence, the interaction term, *CROWDRA* x *NUMFRA*, between *CROWDRA* and *NUMFRA* explains how the impact of *CROWDRA* changes when a user observes her friends rating(s). We also include a user's sequential order (*RSEQ)* in a movie rating queue. For example, a user's *RSEQ* is 1 if she is the first reviewer for a movie. The second user's *RSEQ* is 2, and so on. To compare the effects of *CROWDRA* and *FRWOM* directly, we perform a separate estimation using only the 22% observations in our sample dataset (6,845 observations) in which all users have prior friend ratings.

Alternative product quality information such as marketing effort, box-office sales, average critic rating, and weekly number of reviews are included. Rating users could be already informed or continuously exposed to some degree of quality by advertising. Therefore, cumulative advertising spending (*ADSPEND*) is more relevant than weekly spending. Gross box-office sales for a movie (*TOTBOXSALES*) at a given week are used for movie performance. Weekly number of reviews (*NR*) is included to control for a movie's surrounding popularity level at a given week. Average critic rating (*CR*) is included as one of objective quality information. *MPAA RATING Dummies* are used to control for audience restriction. However, movie dummies are only used when factor loadings are estimated to see any movie specific weight on user rating. As shown similarly in Dellarocas et al. (2006), the correlations among *CROWDRA*, *FRWOM*, and *CR* in our dataset are low.[1] Variance Information Factor (VIF) is less than 3 in our models, therefore multicollinearity is not an issue.

---

[1] Correlations between *CROWDRA* and *CR* = 0.22, between *FRWOM* and *CR* = 0.09, and between *CROWDRA* and *FRWOM* = 0.34 when *FRWOM* is not zero.

Our independent variable set includes several important variables such as an online user's partial demographic feature, online characteristics, visibility and partial social networks in Flixster, e.g., *Gender* and *AGE* , membership duration (*MEMFOR*), the number of generated ratings (*NUMR*) and text-reviews (*NUMRE*), the number of profile viewed by others (*PRFV*), the number of friends (*NUMF*). Table 4 summarizes the independent variables.

4. Observational Learning from Other Users' Movie Ratings

Ideally, our dependent variable would be a continuous value of a user's satisfaction or recommendation about a movie. However, such observation is unavailable. Instead, we consider observed user ratings in item response approach with latent variables. This is appropriate since in our case, rating is a user's decision making of choosing one among ten values for each movie. Latent variables can represent continuous variables underlying observed 'coarsened' responses such as dichotomous or ordinal responses (Skrondal and Rabe-Hesketh 2004). A user can rate multiple movies. Individual ratings are nested in a movie's playing weeks since the movie is released. This structure of user movie ratings over time creates several benefits to our study. First, we can manage the differences between movies. Second, we can explain the variability in the response variable (rating) in terms of variability in observed covariates and unobserved heterogeneity at individual level. The third level, time of movie showing week, gives us an idea about how rating behavior would change over time based on a variety of movie quality information.

Hence, our first model to estimate the impact of the crowd and friend ratings on user movie rating is grounded on an ordered logistic model to relate latent observations to user's ordinal response variable (rating category). Let the latent response $R^*_{i,j,t}$ be a true rating for movie i by reviewer j at time t for the error-prone observed rating $R_{i,j,t}$ due to various noise factors that a reviewer has and the restrictive scale of ratings. We model the true rating as: $R^*_{i,j,t} = \beta X_{i,j} + \gamma Z_{i,j,t} + \delta M_{i,t} + \varepsilon_{i,j,t}$. X is a set of individual user specific variables such as gender, age, and online profiles in Flixster. Z is a set of precedent users' rating information, average ratings of the crowd and friends, and the volume of friend ratings. M is a set of movie specific variables such as advertising spending and performance. Let s be a rating category (from 1 to 10). The observed rating responses are generated by applying thresholds κ_s as follows:

$$R_{i,j,t} = \begin{cases} 1 & \text{if } R^*_{i,j,t} \leq \kappa_1 \\ 2 & \text{if } \kappa_1 < R^*_{i,j,t} \leq \kappa_2 \\ \vdots & \vdots \\ 10 & \text{if } \kappa_9 < R^*_{i,j,t} \end{cases}$$

Here we assume that κ_s, $s = 1, \ldots, S$ are the same for all movies.[2]

Under the assumptions that the error $\varepsilon_{i,j,t}$ has an extreme value distribution and the coefficients are the same for all values of s (*proportional odds assumption*), we model the probability that user j chooses her rating smaller than or equal to s for movie i at time t is

$$P(R_{i,j,t} \le s) = P(R^*_{i,j,t} \le \kappa_s) == \frac{\exp(\kappa_s - \beta X_j - \gamma Z_{i,j,t} - \delta M_{i,t})}{1 + [\exp(\kappa_s - \beta X_j - \gamma Z_{i,j,t} - \delta M_{i,t})]}, s = 1, 2, \ldots, 9.$$

This parallel model becomes equivalent to a series of binary logistic regressions where rating categories are combined (e.g. for $s = 1$ category, rating value 1 is contrasted with rating from 2 to 10; for $s = 2$, the contrast is between rating 1 and 2 versus rating from 3 to 10; and for $s = 9$, rating from 1 to 9 versus rating 10. Williams 2006). Specifically, the true user rating $R^*_{i,j,t}$ can be written as

$$
\begin{aligned}
R^*_{i,j,t} = {} & \beta_1 GENDER_j + \beta_2 AGE_j + \beta_3 Log[MEMFOR_j] + \beta_4 Log[PRFV_j] \\
& + \beta_5 Log[NUMF_j] + \beta_6 Log[NUMRA_j] + \beta_7 Log[NUMRE_j] + \beta_8 RSEQ_{i,j,t} \\
& + \gamma_1 CROWDRA_{i,j,t} + \gamma_2 CROWDRA \times NUMFRA_{i,j,t} + \gamma_3 FRWOM_{i,j,t} \\
& + \delta_1 Log[ADSPEND_{i,t}] + \delta_2 Log[ADSPEND_{i,t}]^2 + \delta_3 Log[NR_{i,t}] + \delta_4 Log[CR_i] \\
& + \delta_5 Log[TOTBOXSALES_{i,t}] + \delta_6 MPAADUMMIES_i + \delta_8 WEEKS_{i,t} + \varepsilon_{i,j,t}
\end{aligned} \tag{1}
$$

We first assume that *CROWDRA*, *NUMFRA*, *FRWOM* and public information about a movie may directly affect a user's rating response. Table 5 presents rating distribution for the sample of 17 movies.

Table 5. User Rating Distribution for 17 movies

	Frequency	Percentage
1	386	1.26%
2	356	1.16%
3	526	1.72%
4	953	3.11%
5	1,435	4.69%
6	3,154	10.3%
7	4,077	13.32%
8	5,812	18.98%
9	5,172	16.89%
10	8,743	28.56%
Total	30614	100%

4.1 User Rating Behavior

The Role of User Characteristics in Movie Rating. Table 6 shows the results of ordered logistic regression estimation using (1) for latent rating response. The column of Model 1-1 in Table 6 runs the regression based on all observations and the column of Model 1-2 is the results with individual rating observations which have friend ratings (*FRWOM*) as well as *CROWDRA*. Therefore, *NUMFRA* in Model 1-1 is an indicator whether a user has at least one prior friend rating.

---

[2] Rating scheme is fixed for all movies and therefore each threshold is homogenous in the sense that reviewers choose the thresholds in the fixed values for every movie - equivalent parameterization (Williams 2006).

Overall, males tend to generate lower ratings than females. Younger users choose higher ratings. More interestingly, the intensity of a user's online activity in Flixster is negatively related to her rating response. For example, if a user has longer membership duration (*MEMFOR*) or a larger number of rating (*NUMRA*) or text review (*NUMRE*) history in Flixster, she is more likely to choose a lower rating (the estimates of β_1, β_2, β_3, β_6, and β_7 are negative and significant at 0.1% level in Model 1-1). In contrast, a user's visibility measured by the number of her profile page viewed by others (*PRFV*) and the number of friends in Flixster (*NUMF*) increases the likelihood of choosing higher rating categories (the estimates of β_4 and β_5 are positive and significant at 0.1% level in Model 1-1). A later user in a movie's rating sequential queue tends to lower her rating (the estimate of $\beta_8 < 0$). The results in Model 1-2 column using only user rating observations which has prior friend ratings show the similar behavior except for *PRFV*.

Table 6. Ordered Logistic Regression (1)

Variables	Model 1-1		Model 1-2	
β_1 [GENDER]	-0.535***	(0.021)	-0.586***	(0.046)
β_2 [AGE]	-0.014***	(0.001)	-0.011***	(0.003)
β_3 LOG[MEMFOR]	-0.333***	(0.048)	-0.448***	(0.100)
β_4 LOG [PRFV]	0.069***	(0.001)	-0.056*	(0.023)
β_5 LOG [NUMF]	0.102***	(0.012)	0.277***	(0.029)
β_6 LOG [NUMRA]	-0.063***	(0.010)	-0.029	(0.021)
β_7 LOG [NUMRE]	-0.125***	(0.009)	-0.135***	(0.022)
β_8 [RSEQ]	-0.002***	(0.000)	0.000	(0.001)
γ_1 [CROWDRA]	1.065***	(0.022)	0.766***	(0.050)
γ_2 [CROWDRA*NUMFRA]	-0.001*	(0.001)	0.000	(0.001)
γ_3 [FRWOM]	-	-	0.289***	(0.015)
δ_1 LOG[ADSPEND]	-0.203***	(0.055)	0.029	(0.115)
δ_2 LOG [ADSPEND^2]	0.053	(0.059)	0.041	(0.135)
δ_3 LOG[NR]	-0.086***	(0.024)	-0.025	(0.047)
δ_4 [CR]	-0.064***	(0.010)	-0.003	(0.021)
δ_5 LOG[TOTBOXSALES]	0.097**	(0.033)	-0.023	(0.069)
δ_6 [RD]	0.125**	(0.046)	0.100	(0.107)
δ_7 [PG13D]	0.058	(0.036)	-0.006	(0.091)
δ_8 [WEEKS]	0.006	(0.007)	-0.001	(0.013)
Log-likelihood	-55871.63		-12099.30	
Number of obs.	30614		6845	
VIF	2.75		2.52	

Note. Standard errors in parentheses. ***p<0.001, **p<0.01, *p<0.05.
The estimates of thresholds (k's) omitted but are available upon request from the authors.

The Role of Movie Specific Information in Movie Rating. δ's in Table 6 are the estimated parameters of movie specific variables. Except for *TOTBOXSALES* and *RD* (*Dummy for an R-rated movie*), other movie specific variables in Model 1-1 are negatively related with a user rating. This may indicate that the higher values of advertising spending, weekly number of reviews, and average critics

rating (*CR*) for a movie can lead a user to choose a lower rating category. This demonstrates that a user may be somewhat disgruntled after seeing a movie when the movie's public quality measures become greater.

Observational Learning in Movie Rating. In our model, observational learning in user movie rating is measured by three factors (e.g. estimated γ's in Table 6): *CROWDRA*, *FRWOM* and *NUMFRA*. If everything else is fixed and these variables are orthogonal to the error term ($\varepsilon_{i,j,t}$), in the results of Model 1-1 in Table 6, we find that a one unit change in the crowd rating (*CROWDRA*) results in three times as large the odds of choosing higher rating categories than lower rating categories (the estimate of $\gamma_2=1.065$ and at 0.1% significant level).[3] By observing a higher average user rating only, a user is more likely to choose a higher rating than a lower rating. The effect is positive and significant in Model 1-2 column in Table 6. However, the effect is relatively smaller when a user can see at least one prior rating by her friends (the estimate of $\gamma_2=0.766$ and at 0.1% significant level).

Comparable Effects of Crowd and Friend Ratings. In contrast to the effect of average user rating, we find that the comparable effect of friend rating becomes much smaller. Estimates of γ_1 and γ_3 in Model 1-2 in Table 6 are positive and significant (the estimates are 0.766 and 0.289 respectively at 0.1% significant level). The estimates are directly comparable since the scales of *CROWDRA* and *FRWOM* are the same. The difference is mainly attributed to intrinsic social interactions embedded in ratings by others, as described earlier. Furthermore, the volume of friend ratings (*NUMFRA*) that a user can observe moderates the effect of crowd rating in Model 1-1 in Table 6. *NUMFRA* not only indicates the presence of friend rating but also measures the amount. Therefore, the estimate of γ_1 (-0.001 and significant at 5% level) in Model 1-1 in Table 6 shows that herding behavior becomes weaker with respect to the volume of friend ratings. This is the evidence of moderating effect by social interactions when we consider a social network among users within a movie boundary. For instance, *NUMFRA* is an in-degree measure (Wasserman and Faust 2007) for a user within a movie's social network in our study.

In contrast, a user's total number of friends[4] (*NUMF*) in Flixster has a positive effect on her rating response consistently across different models in our analysis (e.g., the estimate of $\beta_5=0.102$ in Model 1-1 and 0.277 in Model 2-1 at 0.1% significant level, and similar results can be observed without the parallel line assumption). The result still holds when we test the models again including the interaction term between *CROWDRA* and *NUMF*. Hence, visibility and social interactions of users in a movie specific social network demonstrate different roles in terms of observational learning in our study.

---

[3] The effect over movie running weeks is almost the same and highly significant when we run the model within each movie screening week.

[4] *NUMF* is a proxy for a user's degree centrality in the entire social network in Flixster while *NUMFRA* is considered as in-degree measure in a sub-network for a specific movie in Flixster.

Varying Effects of Independent Variables. We re-estimated the models using generalized ordered logistic regression, which relaxes the parallel line assumption of ordered logistic regression model.[5] Therefore, it allows the parameters to vary according to rating categories by a series of binary logistic regressions. Brant test [6] indicates that some user characteristic variables violate the parallel line assumption. The strongest negative effects of gender, age and membership duration are found with the most extreme review attitude. In other words, male's (female's) negative (positive) rating attitudes become greater when they contrast the rating values below 9 with above 9 (or 1 with above 1). However, the strongest negative effect of rating history is found in the extremely positive rating attitude.

The effect of *CR* becomes positive or negative depending on rating categories (See Figure 4). For example, a higher *CR* leads a user to choose lower rating when she contrasts extremely positive rating with other ratings (e.g., contrasting above 9 with below 9, 1 to 8 versus 9 and 10, and 1 to 9 versus 10). However, there is a positive effect of *CR* when a user contrasts extremely negative rating with other ratings (e.g., contrasting 1 with above 10). Figure 4 depicts how the effect changes according to the rating categories (therefore, H3A and H3B are not supported).
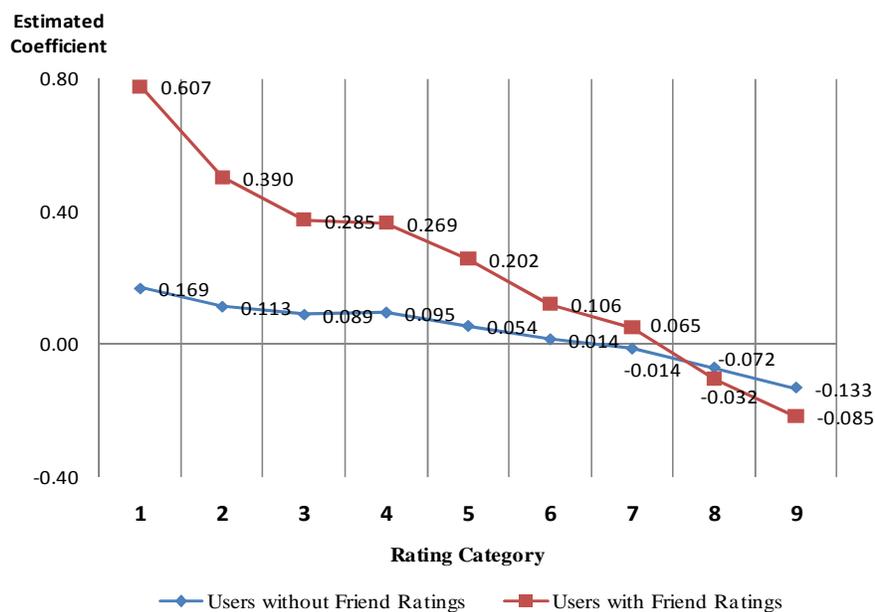
Figure 4. Varying Effects of Critic Rating on User Rating

---

[5] Non-parallel line assumption fit the models in order to verify the effects of the variables are different across rating categories in our models. Therefore, it also test whether there is excess variability between rating categories (Williams 2006)

[6] After testing parallel line assumption using the 5% level of significance, we find that the estimated coefficients of *GENDER*, *AGE*, *MEMFOR*, *NUMRA*, and *NUMRE* vary between rating categories. The testing results and estimates for these variables are omitted to save space but are available from the authors upon request.

The estimated parameters of γ_1 and γ_3 under non-parallel line assumption are shown in Figure 5 graphically. The effect of crowd rating becomes stronger when a user contrasts rating values in extremely favorable rating categories. Hence, a higher average user rating may drive a subsequent user to choose a higher rating category even more strongly when she contrasts ratings in a positive value range such as from 8 to 10. In Table 7, the moderating effect (γ_2) by social interactions becomes only significant when a user contrast relatively neutral rating values (e.g., rating 6 and 7). Therefore, the effect may rely on which rating categories a user contrasts even though the overall effect is negative in Table 6. Also, generalized ordered logistic regression fits the models better according to the improved log-likelihood from Table 6 to Table 7.
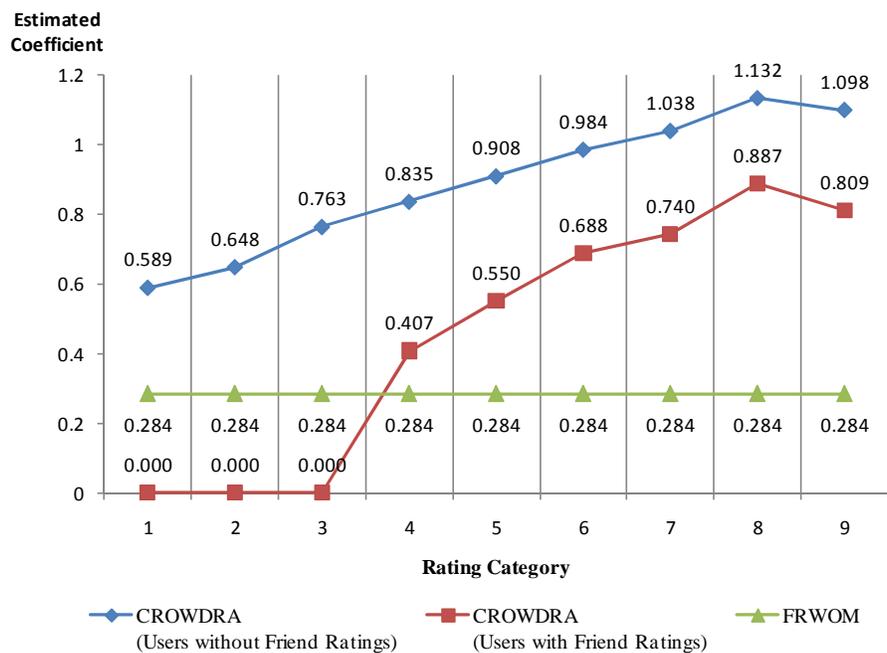
Figure 5. Varying Effects of CROWDRA and FRWOM on User Rating

4.2. Heterogeneity and Self-Perceived Quality

One major issue in our estimation is the unobserved individual heterogeneity. Individuals differ from each other in many aspects. Their ways of thinking, expressing, and participating in generating reviews would be attributed to their own tastes and preferences. Some relevant variables in the user level may not be observed and this leads to unobserved heterogeneity. Without considering this, explaining online user rating behavior on other factors can be biased and misleading. Hence, the latent feature of response and hypothetical development of unobserved variables such as individual self-perceived quality and heterogeneity using latent variables attempts to address this issue.

Table 7. Generalized Ordered Logistic Regression of Model (1)

	Model 1-1		Model 1-2	
Rating 1				
γ_1 [CROWDRA]	0.589***	(0.092)	-0.186	(0.233)
γ_2 [CROWDRA*NUMFRA]	0.008	(0.005)	0.011	(0.008)
γ_3 [FRWOM]	-	-	0.284***	(0.015)
Rating 2				
γ_1 [CROWDRA]	0.648***	(0.068)	-0.243	(0.166)
γ_2 [CROWDRA*NUMFRA]	0.005	(0.003)	0.003	(0.004)
γ_3 [FRWOM]	-	-	0.284***	(0.015)
Rating 3				
γ_1 [CROWDRA]	0.763***	(0.054)	0.160	(0.122)
γ_2 [CROWDRA*NUMFRA]	0.002	(0.002)	0.002	(0.002)
γ_3 [FRWOM]	-	-	0.284***	(0.015)
Rating 4				
γ_1 [CROWDRA]	0.835***	(0.043)	0.407***	(0.093)
γ_2 [CROWDRA*NUMFRA]	-0.001	(0.001)	-0.002	(0.001)
γ_3 [FRWOM]	-	-	0.284***	(0.015)
Rating 5				
γ_1 [CROWDRA]	0.908***	(0.035)	0.550***	(0.076)
γ_2 [CROWDRA*NUMFRA]	-0.001	(0.001)	-0.001	(0.001)
γ_3 [FRWOM]	-	-	0.284***	(0.015)
Rating 6				
γ_1 [CROWDRA]	0.984***	(0.029)	0.688***	(0.063)
γ_2 [CROWDRA*NUMFRA]	-0.002**	(0.001)	-0.002*	(0.001)
γ_3 [FRWOM]	-	-	0.284***	(0.015)
Rating 7				
γ_1 [CROWDRA]	1.038***	(0.026)	0.740***	(0.057)
γ_2 [CROWDRA*NUMFRA]	-0.002***	(0.005)	-0.002*	(0.001)
γ_3 [FRWOM]	-	-	0.284***	(0.015)
Rating 8				
γ_1 [CROWDRA]	1.132***	(0.025)	0.887***	(0.058)
γ_2 [CROWDRA*NUMFRA]	0.000	(0.001)	0.001	(0.001)
γ_3 [FRWOM]	-	-	0.284***	(0.015)
Rating 9				
γ_1 [CROWDRA]	1.098***	(0.028)	0.809***	(0.065)
γ_2 [CROWDRA*NUMFRA]	-0.001	(0.001)	0.000	(0.001)
γ_3 [FRWOM]	-	-	0.284***	(0.015)
Number of Obs.	30614		6845	
Log-Likelihood	-55528.26		-11955.07	

Note.
Standard errors in parentheses. ***p<0.001, **p<0.01, *p<0.05. VIF = 2.75
All other estimates omitted but are available upon request from the authors.

Since causal processes operate at the individual level and not the aggregate level of user ratings, it follows that investigation of causality requires individual-specific effects (Rabe-Hesketh et al. 2004a). From (1) with the assumption of direct impact of other users' ratings, we include two random intercepts to explain individual heterogeneity and time related variation. This yields a model with three different levels

of movie, individual user, and time component. Therefore user j's rating response for movie i in movie screening week t is modeled as:

$$R_{i,j,t}^* = \beta X_j + \gamma Z_{i,j,t} + \delta M_{i,t} + \zeta_j^{(2)} + \zeta_t^{(3)} + \varepsilon_{i,j,t}, \qquad (2)$$

where $\zeta_j^{(2)}$ is an individual-level random intercept, $\zeta_t^{(3)}$ is a movie screening week specific random intercept, and $\varepsilon_{i,j,t}$ has a logistic distribution. We further assume $\zeta_j^{(2)} \sim N(0, \psi^{(2)})$ and $\zeta_t^{(3)} \sim N(0, \psi^{(3)})$.

As mentioned earlier, we can only observe reviewer's rating in a fixed scale and therefore observed rating, $R_{i,j,t}$, are generated by the threshold model, and we assume that κ_s's ($s = 1, \ldots, 10$) are the same for all movies.[7] We implement generalized linear latent variable and mixed models (GLLAMM)[8] to estimate our parameters. This allows maximizing the likelihood of the conditional density of the response variable given the latent and explanatory variables with the prior density of the latent variables with adaptive quadrature (see the appendix for detailed estimation procedure).

Table 8. Regression Results for the Variances of Random Intercepts

Parameters	Single Level		Two Level		Three Level	
	Est.	(SE)	Est.	(SE)	Est.	(SE)
User-level $\psi^{(2)}$	-	-	2.356***	(0.104)	2.355***	(0.214)
Time-level $\psi^{(3)}$	-	-	-	-	0.004	(0.003)
Thresholds						
κ_1	-4.467	(0.057)	-5.640	(0.074)	-5.625	(0.114)
κ_2	-3.799	(0.044)	-4.917	(0.063)	-4.894	(0.102)
κ_3	-3.241	(0.037)	-4.292	(0.056)	-4.259	(0.092)
κ_4	-2.637	(0.032)	-3.589	(0.050)	-3.542	(0.081)
κ_5	-2.070	(0.030)	-2.899	(0.045)	-2.836	(0.070)
κ_6	-1.287	(0.027)	-1.894	(0.040)	-1.810	(0.055)
κ_7	-0.582	(0.026)	-0.945	(0.036)	-0.843	(0.044)
κ_8	0.266	(0.026)	0.232	(0.034)	0.346	(0.040)
κ_9	1.062	(0.027)	1.341	(0.036)	1.457	(0.050)
Log-Likelihood	-56679.912		-55768.037		-56513.934	

Note. Standard errors in parentheses. ***p<0.001, **p<0.01, *p<0.05.
All other estimates omitted but are available upon request from the authors

Empirical Evidence of Unobserved Heterogeneity. In order to verify the unobserved heterogeneity in each level, between users and between weeks, first, we estimate the variances of random intercepts of users and weeks, $\psi^{(2)}$ and $\psi^{(3)}$ respectively with a simpler model (2) which only includes the interaction terms of movie dummies and *CROWDRA*. Table 8 shows the results of three different specifications. Single-level represents a single-level ordinal response model without any random effect; two-level

---

[7] Rating scheme is fixed for all movies and therefore each threshold is homogenous in the sense that users choose the thresholds in fixed values for every movie.

[8] GLLAMMs are a class of multilevel latent variable models for (multivariate) responses of mixed type including continuous responses, counts, duration/survival data, dichotomous, ordered and unordered categorical responses and rankings (see Skrondal and Rabe-Hesketh 2004).

includes only user specific random intercept; and three-level contains both random intercepts of users and time component. In addition to the variances estimated for the clusters, the estimates of 9 cut-points, κ_1, …, κ_9, appear at the bottom of Table 8.

Table 9. Generalized Latent Linear Mixed Regression (2)

Variables	Model 2-1		Model 2-2	
β_1 [GENDER]	-0.649***	(0.021)	-0.732***	(0.046)
β_2 [AGE]	-0.020***	(0.001)	-0.021***	(0.003)
β_3 LOG[MEMFOR]	-0.350***	(0.048)	-0.527***	(0.100)
β_4 LOG [PRFV]	0.123***	(0.001)	0.009	(0.023)
β_5 LOG [NUMF]	0.105***	(0.012)	0.317***	(0.029)
β_6 LOG [NUMRA]	-0.099***	(0.010)	-0.083*	(0.021)
β_7 LOG [NUMRE]	-0.140***	(0.009)	-0.192***	(0.022)
β_8 [RSEQ]	-0.003***	(0.000)	-0.002	(0.001)
γ_1[CROWDRA]	1.386***	(0.022)	1.050***	(0.050)
γ_2 [CROWDRA*NUMFRA]	-0.001	(0.001)	0.000	(0.001)
γ_3 [FRWOM]	-	-	0.318***	(0.015)
δ_1 LOG[ADSPEND]	-0.247***	(0.055)	0.000	(0.115)
δ_2 LOG [ADSPEND^2]	0.079	(0.059)	0.053	(0.135)
δ_3 LOG[NR]	-0.093***	(0.024)	-0.010	(0.047)
δ_4 [CR]	-0.079***	(0.010)	0.001	(0.021)
δ_5 LOG[TOTBOXSALES]	0.143***	(0.033)	0.043	(0.069)
δ_6 [RD]	0.235***	(0.046)	0.234⁻	(0.107)
δ_7 [PG13D]	0.097*	(0.036)	0.007	(0.091)
δ_8 [WEEKS]	0.012	(0.007)	0.001	(0.013)
Variance ($\psi^{(2)}$)	0.701***	(0.158)	0.634⁻	(0.326)
Factor Loadings				
λ_1[Movie1]	1		1	(Fixed)
λ_2[Movie2]	1.444***	(0.183)	2.001***	(0.546)
λ_3[Movie3]	1.532***	(0.181)	1.618***	(0.432)
λ_4[Movie4]	1.259***	(0.188)	0.832***	(0.350)
λ_5[Movie5]	1.480***	(0.187)	1.622***	(0.477)
λ_6[Movie6]	1.265***	(0.182)	1.354***	(0.412)
λ_7[Movie7]	2.063***	(0.241)	2.323***	(0.612)
λ_8[Movie8]	2.026***	(0.257)	2.379***	(0.692)
λ_9[Movie9]	2.217***	(0.268)	2.613***	(0.724)
λ_{10}[Movie10]	1.912***	(0.255)	1.769***	(0.566)
λ_{11}[Movie11]	1.246***	(0.161)	1.444***	(0.396)
λ_{12}[Movie12]	1.191***	(0.171)	0.551⁻	(0.285)
λ_{13}[Movie13]	2.060***	(0.234)	2.362***	(0.614)
λ_{14}[Movie14]	1.454***	(0.184)	1.623***	(0.446)
λ_{15}[Movie15]	1.847***	(0.211)	2.215***	(0.576)
λ_{16}[Movie16]	1.317***	(0.192)	1.402***	(0.422)
λ_{17}[Movie17]	2.010***	(0.234)	2.529***	(0.653)
Log-likelihood	-55060.10		-11835.80	
Number of obs.	30614		6845	

Note. Standard errors in parentheses. ***$p<0.001$, **$p<0.01$, *$p<0.05$., ⁻$p<0.1$
The estimates of thresholds (κ's) omitted but are available upon request from the authors.

The two-level model shows the unobserved heterogeneity between users with a variance estimated as 2.356 which is significant at 0.1% level. Furthermore, the log-likelihood of the two-level model increases from the single-level. However, the estimated variance component for time is nearly zero and does not appear to be significant in three-level specification,[9] and the log-likelihood of this specification decreases. Therefore, we conclude that we should consider the unobserved heterogeneity in users in our further estimation but the random effect of time component can be ignorable. Nevertheless, we keep a time variable (*WEEKS*) in the models in order to keep track of the direct effect of movie screening week.

Hence, our model (2) accounting for the user level heterogeneity becomes

$$R_{i,j,t}^{*} = \beta X_j + \gamma Z_{i,j,t} + \delta M_{i,t} + \lambda_i \zeta_j^{(2)} + \varepsilon_{i,j,t}$$

and we add factor loading (λ_i) to account for movie specific difference in rating response as well. Table 9 reports the estimation results of two different specifications. Both models in Table 9 fit significantly better than the corresponding single level models in Table 6 based on log-likelihood. The variance of user specific random intercept is significant in both models in Table 9 (the estimate of $\psi^{(2)} = 0.701$ at 0.1% significant level in Model 2-1 and 0.634 in Model 2-2 at 10% significant level).

The results in Table 9 are consistent with those in Table 8 with respect to the signs. In addition, the estimated factor loadings[10] report users' relative weight on a specific movie in rating response. For example, Movie 9 (*Ghost Rider*) has the greatest factor loading (λ_9). Interestingly, this movie is the lowest rated movie by critics among our sample movies. It can be observed that the factor loading becomes relatively greater as the gap between average user rating and critics rating increases in Table 2. This may explain users' opposite reaction in rating response to critic rating for some movies. We find a stronger effect of *CROWDRA* than *FRWOM* on user rating, similarly as before. However, the moderating effect (*CROWDRA* x *NUMFRA*) becomes insignificant in the column of Model 2-1 and we examine this again after accounting for other issues.

Self-Perceived Quality. A user's self-perceived quality after watching a movie cannot be observed but may be influenced by others' online opinions. Since the perceived quality can be related to user rating response, we construct a hypothetical variable for self-perceived quality of user *j* in (2):

$$R_{i,j,t}^{*} = \beta X_j + \delta M_{i,t} + \lambda_i \eta_{i,j}^{(2)} + \varepsilon_{i,j,t} \quad (3)$$

where $\eta_{i,j}^{(2)}$ is user *j*'s self-perceived quality for movie *i* at time *t* and λ_i is a factor loading for movie *i*, and it can be modeled as a structural equation as,

$$\eta_{i,j}^{(2)} = \gamma Z_{i,j,t} + \zeta_j^{(2)} \quad (3.1)$$

---

[9] We also re-run model (2) in Table 8 using different variable sets and the results are not significantly different.

[10] The scale of the factor is identified through the constraint that the first factor loading (λ_1) equals 1.

where $\zeta^{(2)}{}_j$ is an individual level random intercept and $Z_{i,j,t}$ contains social influence variables: *CROWDRA*, *FRWOM*, and *NUMFRA*. Hence, we examine how ratings by others affect users' rating responses *indirectly*. The estimation procedure is the same as that of (2), and the results are reported in Table 9-1.

This model assumes that others' ratings affect a subsequent users' rating only through the latent perceived quality $\eta_{i,j}^{(2)}$, i.e., the effect of *CROWDRA* on the movie i is to increase the latent rating response by $0.482*\lambda_i$, and the effect of *CROWDRA* x *NUMFRA* is to reduce by $-0.001*0.482*\lambda_i$, in Model 3-1 of Table 9-1. Hence, the herding behavior and moderating effect by the volume of friend ratings still hold when we assume that observational learning and social interactions indirectly affect a user' rating response through unobserved factors such as user perceived quality. The effects vary according to factor loading (λ_i). In Model 3-2 of Table 9-1, we can compare directly the effects of two types of observational ratings by others (*CROWDRA* and *FRWOM*) for only those who have both observational sources. Interestingly, the relative effect of *FRWOM* becomes greater than *CROWDRA* even though both effects are still positive and significant. Therefore, we conclude that users may weigh more on *FRWOM* than *CROWDRA* when they update their private quality measures which indirectly affect their rating responses.

Table 9-1. GLLAMM with a Structural Eq. for Perceived Quality (3)

Variables	Model 3-1		Model 3-2	
	CROWDRA	CROWDRA *NUMFRA	CROWDRA	FRWOM
Indirect Effect	$\lambda_{i*}\gamma_1$	$\lambda_{i*}\gamma_1$	$\lambda_{i*}\gamma_1$	$\lambda_{i*}\gamma_1$
Movie1	0.482	-0.001	0.121	0.243
Movie2	0.565	-0.002	0.152	0.306
Movie3	0.539	-0.002	0.139	0.279
Movie3	0.386	-0.001	0.089	0.178
Movie3	0.487	-0.001	0.115	0.230
Movie3	0.467	-0.001	0.118	0.236
Movie7	0.723	-0.002	0.197	0.396
Movie8	0.542	-0.002	0.152	0.304
Movie9	0.695	-0.002	0.174	0.349
Movie10	0.469	-0.001	0.141	0.283
Movie11	0.498	-0.001	0.125	0.250
Movie12	0.518	-0.002	0.131	0.263
Movie13	0.727	-0.002	0.196	0.394
Movie14	0.462	-0.001	0.114	0.229
Movie15	0.567	-0.002	0.146	0.294
Movie16	0.427	-0.001	0.100	0.200
Movie17	0.672	-0.002	0.187	0.375
Variance ($\psi^{(2)}$)	1.392		1.441	
Log-likelihood	-55075.90		-11848.60	
Number of obs.	30614		6845	

Note. All estimates are significant at less than 5% level.
All other estimates omitted but are available upon request from the authors.

4.3. Reflection Problem

The other key issue in studies attempting to identify the effect of social influence is the reflection problem (Manski 1993). The reflection problem in our context would mean that users choose similar ratings for a movie not because of the influence of observing others' rating outcomes but that she is in the same reference group with others. Users tend to behave similarly because they are alike or face a common environment. Thus, the relationship between observational learning and individual rating outcome could be spurious. The parameter estimates for observational learning variables are biased by endogeneity stemming from this problem.

Following Bramoullé et al (2009) in which they suggest the ways by which the true effect of social influence can be identified by accounting for the reflection problem, we first introduce unobservable variables common to the individuals that belong to the same social network structure for a movie. A network specific unobservable, α_i, captures unobserved variables that have common effects on the rating outcomes of all users within movie i's specific network (e.g., individuals' similar preferences of watching and reviewing the movie). Then, user j's rating response model becomes $R_{i,j} = \alpha_i + \beta X_{i,j} + \gamma Z_{i,j} + \delta M_i + \varepsilon_{i,j}$. The other user $j-1$ has a same rating equation if she belongs to the same social network for movie i, explicitly, $R_{i,j-1} = \alpha_i + \beta X_{i,j-1} + \gamma Z_{i,j-1} + \delta M_{i,j-1} + \varepsilon_{i,j-1}$. Differencing these two equations gives:

$$\Delta R_{i,j} = \beta \Delta X_{i,j} + \gamma \Delta Z_{i,j} + \delta \Delta M_{i,j} + \Delta \varepsilon_{i,j}, \quad (4)$$

where $\Delta y_{i,j} = y_{i,j} - y_{i,j-1}$, for y = X, Z, M. Now, network fixed effects (α_i) potentially correlated with observational influence variables (Z) is cancelled out. Therefore, the model generates internal conditions that ensure identification of social effects in spite of serial correlation of $\Delta \varepsilon_{i,j}$ (Bramoullé et al. 2009). However, the differencing also excludes movie specific variable set, M, due to multicollinearity and time-invariance. After we plot $\Delta R_{i,j}$, it appears normally distributed.

We run (4) by OLS for model specification (1), GLLAM for model specification (2), and GLLAM for model specification (3) and γ's are estimated in Table 10 with other parameters. The column of OLS (1)[11] in Model 4-1 in Table 10 reports the results of our initial model (1). The effect of *CROWDRA* becomes slightly smaller than that in previous analyses but is still significant. The moderating effect by social interactions (*CROWDRA* x *NUMFRA*) becomes greater after accounting for potential movie specific homophily. All other estimates remain the same except the effect of *RESQ* which becomes insignificant. Similarly, the previous results for the effects of individual specific variables hold in

---

[11] We have run a Tobit model with censoring minimum and maximum outcomes (e.g., -9 and 9). The results are qualitatively similar.

GLLAM (2) and (3) in Model 4-1 in Table 10. The direct moderation effect, γ_2, in GLLAM (2) is significant after accounting for individual heterogeneity. The indirect effect of *CROWDRA* still exists but the magnitude decreases greatly, and the indirect effect of *CROWDRA* x *NUMFRA* becomes insignificant in GLLAM (3).

Most results of Model 4-2, reported in Table 10, are very similar to those of previous analyses. This suggests that the impact of crowd rating is greater than friend rating in terms of observational learning. However, the overall effect of *CROWDRA* on the ratings by users who have at least one friend rating is less than those who have no friend rating. Therefore, the presence of friend rating always moderates herding behavior in online user rating. Interestingly, the indirect effect of crowd rating becomes insignificant and only friend rating indirectly affect a user's rating decision. This might explain that individual perceived quality may increase if her friends like the same movie.

5. Empirical Evidence: Impact of User Rating on Movie Performance

We next investigate whether time varying online average user rating (*CROWDRA*) is correlated to subsequent weekend box-office sales (*WKSALES*): H4A and H4B. It is critical in our study to find the theory of positive impact of user rating on consumers' purchasing decisions still holds in the movie market. That is, consumers can compensate for possibly biased user ratings by observational learning to make rational purchase decisions. Therefore, our result may support the finding that strategic manipulation of user ratings (Dellarocas 2006) in terms of herding behavior in online user rating may be optimal for firms.

Our approach is based on prior work (e.g., Chevalier and Mayzlin 2006, Li and Hitt 2008), except that we consider population, rather than sample, sales of movies in 2007. Importantly, movie ticket price is usually fixed and its temporal impact on sales can be ruled out in our pooled multiple movies.[12] Our direct measure of weekly advertising spending does not need to proxy ongoing promotions of movies. To account for possible nonlinear relationship of advertising, we add its square term (*WeeklyADSPEND*^{2}). We control for weekly number of reviews (*NR*) which may capture other idiosyncratic aspects of movie demand not otherwise covered in our model. Relative performance and competition of a movie in its market can be captured by public ranking information (*RANK*) in that week of *WKSALES*. We include a time-trend variable (*WEEKS*) which captures the number of weeks since release, to ensure that we are not confounding our temporal review measure with a simple time trend. Table 11 provides the description of our measures in sales model.

---

[12] However, μ_i captures any movie fixed effects potentially correlated with average user rating and the number of weekly reviews in our sales model (5).

Table 10. Network Fixed Effect Estimation

Variables	Model 4-1						Model 4-2					
	OLS(1)		GLLAM(2)		GLLAM(3)		OLS(1)		GLLAMM (2)		GLLAM(3)	
β_1 [ΔGENDER]	-0.516***	(0.023)	-0.495***	(0.023)	-0.496***	(0.024)	-0.563***	(0.045)	-0.535***	(0.046)	-0.531***	(0.046)
β_2 [ΔAGE]	-0.014***	(0.001)	-0.015***	(0.001)	-0.015***	(0.001)	-0.008***	(0.002)	-0.011***	(0.002)	-0.010***	(0.002)
β_3 ΔLOG[MEMFOR]	-0.267***	(0.052)	-0.262***	(0.052)	-0.253***	(0.053)	-0.346***	(0.099)	-0.359***	(0.099)	-0.355***	(0.100)
β_4 ΔLOG [PRFV]	0.071***	(0.011)	0.070***	(0.011)	0.073***	(0.011)	-0.046*	(0.022)	-0.032	(0.023)	-0.033	(0.023)
β_5 ΔLOG [NUMF]	0.116***	(0.013)	0.108***	(0.013)	0.107***	(0.014)	0.224***	(0.028)	0.210***	(0.029)	0.220***	(0.029)
β_6 ΔLOG [NUMRA]	-0.067***	(0.011)	-0.060***	(0.011)	-0.060***	(0.011)	0.015	(0.021)	0.004	(0.021)	0.007	(0.021)
β_7 ΔLOG [NUMRE]	-0.104***	(0.010)	-0.103***	(0.010)	-0.105***	(0.010)	-0.128***	(0.022)	-0.128***	(0.022)	-0.130***	(0.022)
β_8 Δ [RSEQ]	-0.000	(0.001)	-0.001	(0.001)	-0.001	(0.001)	0.000	(0.001)	0.000	(0.001)	0.000	(0.001)
γ_1 Δ[CROWDRA]	0.912***	(0.124)	0.868***	(0.091)	-		0.408*	(0.194)	0.512*	(0.261)	-	
γ_2 Δ[CROWDRA*NUMFRA]	-0.002***	(0.001)	-0.002**	(0.001)	-		-0.000	(0.001)	0.000	(0.038)	-	
γ_3 Δ[FRWOM]	-		-		-		0.267***	(0.014)	0.250***	(0.014)	-	
Structural Model												
γ_1 Δ[CROWDRA]	-		-		0.170⁻	(0.087)	-		-		0.230	(0.158)
$\gamma_2\Delta$[CROWDRA*NUMFRA]	-		-		-0.001	(0.000)	-		-		-0.000	(0.045)
$\gamma_3\Delta$[FRWOM]	-		-		-		-		-		0.144**	(0.048)
Variance												
$\psi^{(1)}$	-		6.109***	(0.072)	6.106***	(0.072)	-		4..994***	(0.121)	4.989***	(0.122)
$\psi^{(2)}$	-		0.045	(0.054)	0.049***	(0.058)	-		0.126	(0.193)	0.399	(0.264)
Number of obs.	30583		30583		30583		6827		6827		6827	
Log-Likelihood	-		-73551.55		-73553.61		-		-15804.75		-15803.89	
R-Squared	0.043		-		-		0.116		-		-	

Note. Standard errors in parentheses. ***p<0.001, **p<0.01, *p<0.05. ⁻p<0.10
All movie specific variables are dropped after differencing. The estimates of factor loadings and thresholds omitted and are available upon request from the authors

Overall, this yields the following estimating equation:

$$Log[WKSALES_{i,t}] = \mu_i + \pi_0 + \pi_1 CROWDRA_{i,t} + \pi_2 Log[NR_{i,t}]$$
$$+ \theta_1[WeeklyADSPEND_{i,t}] + \theta_2[WeeklyADSPEND_{i,t}]^2 \qquad (5)$$
$$+ \theta_3 RANK_{i,t} + \theta_4 WEEKS_{i,t} + \varepsilon_{i,t}$$

Our sales model is estimated first using the weekend sales dataset of all movies (269 movies) in 2007 which had at least two screening weeks in theaters and observed advertising spending information. Second, we run the same model for the sample of 17 movies which were used to test observational learning in the previous section. The pooled movie set consists of 2,385 observations for 269 movies and the sampled movie set has 256 observations for 17 movies (See Table 11); both datasets, therefore, are unbalanced panel data. After checking for potential endogeneity stemming from the fixed effects (μ_i) by Hausman specification test (χ^2 statistic = 42.80, degree of freedom = 6, and p<0.000), we run model (5) by fixed effect estimation. This estimation is insensitive to the problems of having inadequate controls for time-invariant difference across movies. However, our testing result for heteroskedasticity using Wald test rejects the constant variance assumption in the model (χ^2 statistic = 310.75, degree of freedom= 269, and p<0.000). To fix this, we have computed robust standard errors to control for heteroskedasticity in fixed effect estimation.

Table 11. Description of Our Measures Used in Sales Model

Measures	Description	All movies			17 movies		
		Mean	Min	Max	Mean	Min	Max
WKSALES$_{i,t}$	Weekend sales for movie i at time t in $thousand	23.78	0.001	1511.17	81.02	0.06	1511.17
CROWDRA$_{i,t}$	Average rating of all reviews posted for movie i since it was release until time t	7.55	2	10	7.92	6.99	9.06
NR$_{i,t}$	Number of movie reviews posted on Flixster.com for movie i since it was released at time t	28.30	1	1608	136.50	5	1608
Weekly ADSPEND$_{i,t}$	Weekly advertising spending for movie i at time t in $million	0.34	0	7.14	0.5	0	6.20
RANK$_{i,t}$	Number of weeks for movie i since it was released	36.80	1	138	20.56	1	90
WEEKS$_{i,t}$	movie i's screening week	8.30	1	44	8.95	1	22

The estimates in Fixed Effect columns for both datasets in Table 12 are consistent with our expectation. However, the estimate for *CROWDRA* in our sample movies dataset becomes insignificant. We have checked for serial correlation using Lagrange Multiplier test and found that our datasets have first-order autocorrelation (F(1, 195) = 496.96, p<0.000). To fix this, we use first-differencing approach and this changes our sales model as:

$$\Delta Log[WKSALES_{i,t}] = \pi_0 + \pi_1 \Delta CROWDRA_{i,t} + \pi_2 \Delta Log[NR_{i,t}]$$
$$+\theta_1 \Delta[WeeklyADSPEND_{i,t}] + \theta_2 \Delta[WeeklyADSPEND_{i,t}]^2 \quad (5\text{-}1)$$
$$+\theta_3 \Delta RANK_{i,t} + \Delta \varepsilon_{i,t}$$

where $\Delta Log[WKSALES_{i,t}] = Log[WKSALES_{i,t}] - Log[WKSALES_{i,t-1}]$ and other variables are generated in the same manner.[13] Then, the results by ordinary least square estimation of (5-1), as presented in Table 12, are unbiased and efficient after accounting for all possible estimation issues described above.

Table 12. Fixed Effect and First Differencing estimations of Weekend Box Office Sales Model

Dependent Variable	All Movies (269)				Sample Movies (17)			
	Fixed Effect		First Differencing		Fixed Effect		First Differencing	
LOG[WKSALES$_{i,t}$]								
Variable								
π_1CROWDRA$_{i,t}$	0.199***	(0.060)	0.185*	(0.074)	0.430	(0.590)	0.959*	(0.406)
π_2LOG[NR$_{i,t}$]	0.291***	(0.015)	0.096***	(0.015)	0.157*	(0.060)	0.118*	(0.049)
θ_1[WeeklyADSPEND$_{i,t}$]	0.786***	(0.031)	0.357***	(0.032)	0.657***	(0.091)	0.302***	(0.057)
θ_2[WeeklyADSPEND$_{i,t}$]^2	-0.110***	(0.008)	-0.038***	(0.008)	-0.078**	(0.022)	-0.025**	(0.009)
θ_3RANK$_{i,t}$	-0.061***	(0.001)	-0.061***	(0.001)	-0.058***	(0.007)	-0.054***	(0.007)
θ_4WEEKS$_{i,t}$	-0.043***	(0.003)	-		-0.123***	(0.026)	-	
π_0	1.172**	(0.456)	-0.116***	(0.021)	0.377	(4.582)	-0.170***	(0.037)
Number of Obs.	2385		2116		256		239	
R^2-overall	0.9506		0.7429		0.9580		0.531	
Auto Correlation Test	F(1, 195) = 496.96		F(1, 184) = 0.589		F(1, 195) = 24.89		F(1, 184) = 2.097	
Hausman Test	χ^2(6) = 42.80		-		χ^2(6) = 10.17		-	
Heteroskedasticity Test	χ^2(269) = 310.75		-		χ^2(17) = 123.20		-	
VIF	2.73		2.20		5.04		1.99	

Note. Robust Standard Errors in parentheses; *p<0.05, **p<0.01, ***p<0.001

On average, weekend box-office sales decreases with *RANK* and *WEEKS,* and increases with the number of reviews (*NR*) and advertising (*weeklyADSPEND*). Consistent with prior work on online book sales (Chevalier and Mayzlin 2006, Li and Hitt 2008), average user rating (*CROWDRA*) is positively correlated with sales (π_1=0.185, p<0.05). This supports H4A of the positive impact of user rating on weekend box-office sales (H4B is supported for popular movie sets as well in Table 12, π_1=0.959, p<0.05). For example, a 0.5 (in a 1 to 10 scale) increase in average user rating can increase weekend box-office sale by 10%, which is close to the result on book sales in Amazon.com (Li and Hitt 2008). More interestingly, a 0.5 increase in *CRWODRA* is equal to the effect of $0.12 million increase in weekly advertising spending to increase weekend box-office sales by 10%. The effect of average user rating becomes even greater with popular movies. A 0.5 increase in average user rating increases the sales by 62%

---

[13] *WEEKS* is dropped since the lag difference created *ΔWEEKS* =1 for all observations.

and is equivalent to $1.6 million increase in weekly advertising. The sample movies generate a relatively large amount of user reviews (see Table 3) and attract much more advertising budget than average movies (e.g., average ad spending after release is $14 million for the sample movies and $7.5 for all movies in 2007). Hence, average user ratings in popular movies are relatively harder to increase across weeks and therefore the relative impact of user rating is much greater on the box office performance of popular movies.

Hence, by supporting H4A and H4B, our results of observational learning in user rating are very significant to firms. If a movie begins to generate higher user ratings in its beginning period of screening in theaters, its average user rating is more likely to move towards positive direction rather than negative direction by herding behavior of user ratings. Consequently, this will positively impact long term box office sales.

6. Conclusion and Discussion

This paper studies how online user rating is generated based on observational learning from others' rating decisions and product information. Existing research has focused on the relationship between user reviews and sales. However, understanding the social drivers of a user's rating can help managerial practices such as tailored marketing strategy and reliable design of recommender system.

Our analysis suggests that observational learning by others' ratings can trigger herding behavior when a subsequent generates movie rating. Consequently, aggregate level of consumer-generated product rating information (e.g., average user rating) may not be an unbiased indication of unobserved quality (Li and Hitt 2008). The herding behavior can evolve positively or negatively since the probability that a user choose a high or low rating depends on the valence of prior user reviews. Based on our sales model analysis, positive herding in online user movie rating will increase box office sales, while negative herding may hurts sales. Our analysis shows an equivalent effect of average user rating to advertising spending. Thus, firms can benefit by tailoring their marketing strategies more effectively to take into account this consumer behavior in generating online reviews.

On the other hand, the presence of observational learning in online user rating lowers quality information created by users since each user rating would be associated with some degree of bias due to herding behavior. From theory of informational cascades (Banerjee 1992, Bikhchandani et al. 1992, 1998) and our empirical results, we can provide the ways of investing to alleviate this bias to review systems or recommend system designers to increase the integrity of quality information created by users online. When users generate reviews, minimizing influence by aggregate information such as average user rating or any numerical summary can be conducive to increase quality of online user reviews. Second, maximizing the opportunity for users to share their product experiences can moderate herding behavior. However, information overloads and online anonymity make information flow difficult. Hence,

increasing social interactions among online friends can be an effective solution to flow private quality smoothly in a large-scale consumer network. Increasing the visibility of friend recommendation is another effective way to prevent this bias in online user reviews since our result shows that friend rating is less influenced in terms of observational learning in user rating decision.

This study can be extended in a few directions. If we can capture private information along with user ratings, we would be able to fully explain how user ratings deviate from each other and the underlying reason for that in population and socially local groups. Hence, one possible extension to this paper is to include private information from text review to test whether its role is different with discrete rating in recommendation system. Second, most recommender systems use both numerical rating and text review together at one unit review level. Some people write overall positive attitude for products in text reviews while generating relatively lower product ratings or vice versa. This discrepancy between text review and rating may be attributed to not only individuals' different preferences and but also the influence of other reviews or public product quality information. However, it is not clear how these two review devices are different in terms of the informativeness of product quality and to what extent each of them affects consumer behavior with respect to generating online reviews and product performance. Finally, our model can be extended to examine the influence of online user rating on the DVD market in order to account for the long tail phenomenon (Anderson 2006) in movies.

References

Anderson, C. 2006. *The Long Tail. Why the Future of Business is Selling Less of More*. New York: Hyperion.

Anderson, E.T., D.I. Simester. 2001. Price discrimination as an adverse signal: Why an offer to spread payments may hurt demand. *Marketing Sci*. 20(3) 315–327.

Anderson, L.R., C.A. Holt. 1997. Information cascades in the laboratory. *Amer. Econom. Rev*. 87(5) 847–862.

Arndt, J. 1967. Role of Product-Related Conversations in the Diffusion of a New Product. *Journal of Marketing Research* 4(3) 291-295.

Ba, S., P.A. Pavlou. 2002. Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior. *MIS Quarterly* 26(3) 243-268.

Bajari, P., A. Hortacsu. 2003. The Winner's Curse, Reserve Prices, and Endogenous Entry: Empirical Insights from eBay Auctions. *RAND Journal of Economics* 34(2) 329-355.

Banerjee, A.V. 1992. A simple model of Herd behavior. *Quart. J. Econom* 107(3) 797–817.

Banerjee, A. 1993. The economics of rumors. *Rev. Econ. Stud.* 19 309–327.

Bass, F.M. 1969. A New Product Growth Model for Consumer Durables. *Management Sci.* 15 215-227.

Baumeister, R.F., M.R. Leary. 1995. The need to belong: Desire for interpersonal attachment as a fundamental human motivation. *Psych. Bull.* 117(3) 497.

Bikhchandani, S., D. Hirshleifer, I. Welch. 1992. A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *The Journal of Political Economy* 100(5) 992-1026.

Bramoullé, Y, H. Djebbari, B. Fortin. 2009. Identification of peer effects through social networks. *Journal of Econometrics* 15 41-55.

Brewer, M. 1991. The social self: On being the same and different at the same time. *Personality Soc. Psych. Bull* 17(5) 475–482.

Brynjolfsson, E., M.D. Smith. 2000. Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Science* 46(4) 563-585.

Camerer, C., T. Ho. 1999. Experience-weighted Attraction Learning in Normal Form Games. *Econometrica* 67(4) 827-874.

Çelen, B., S. Kariv. 2004a. Distinguishing Informational Cascades from Herd Behavior in the Laboratory. *American Economic Review* 94(3) 484-497.

Çelen, B., S. Kariv. 2004b. Observational learning under imperfect information. *Games and Economic behavior* 47(1) 72-86.

Chevalier, J.A., D. Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research* 43(3) 345–354.

Chen, P.-Y., S.-Y. Wu, J. Yoon. 2004. The Impact of Online Recommendations and Consumer Feedback on Sales. *Proceedings of the 25th International Conference on Information Systems* 711-724.

Coleman, J.S., E. Katz, H. Menzel. 1966. *Medical Innovation: A diffusion study*. Indianapolis: Bobbs-Merill.

Dellarocas, C. 2003. The digitization of word-of-mouth: Promise and challenges of online reputation mechanisms. *Management Science* 49(10) 1407-1424.

Dellarocas, C. 2006. Strategic manipulation of Internet opinion forums: Implications for consumers and firms. *Management Sci.* 52(10) 1577–1593.

Dellarocas, C., X. Zhang, N.F. Awad. 2007. Exploring the Value of Online Product Ratings in Revenue Forecasting: The Case of Motion Pictures. *Journal of Interactive Marketing* 21(4) 23.

Desai, P.S. 2000. Multiple messages to retain retailers: Signaling new product demand. *Marketing Sci.* 19(4) 381–389.

Dobuzinskis, A. 2009. Flixster's "Movies" aims to connect film fans. *Reuters.com*. Available at http://www.reuters.com/article/idUSTRE57R43T20090828.

Duan, W., B. Gu, A.B. Whinston. 2005. Do online reviews matter?—An empirical investigation of panel data. Working paper, The University of Texas at Austin.

Duan, W., B. Gu, A.B. Whinston. 2009. Informational Cascades and Software Adoption on the Internet: An Empirical Investigation. *MIS Quarterly* 33(1) 23-48.

Ellison, G., D. Fudenberg. 1995. Word-of mouth communication and social learning. *Quarterly Journal of Economics* 110 95–126.

Erdem, T., M. Keane. 1996. Decision Making under Uncertainty: Capturing Dynamic Brand Choice Processes in Turbulent Consumer Goods Markets. *Marketing Science* 15(1) 1-20.

Forman, C., A. Ghose, B. Wiesenfeld. 2008. Examining the Relationship Between Reviews and Sales. *Information Systems Research* 19(3) 291–313.

Galdwell, M. 2000. *The Tipping Point: How little things can make a big difference*. Little, Brown and Company.

Glockner, P. 2009. Nanocrowd Has a New Take on Movie Recommendations. *Readwriteweb.com*. http://www.readwriteweb.com/archives/nanocrowd_has_a_new_take_on_movie_recommendations.php.

Godes, D., D. Mayzlin. 2004. Using online conversations to measure word-of-mouth communication. *Marketing Science* 23(4) 545–560.

Guo, L., Y. Zhao. 2009. Voluntary quality disclosure and market interaction. *Marketing Sci.* 28(3) 488–501.

Hitsch, G.J. 2006. An Empirical Model of Optimal Dynamic Product Launch and Exit Under Demand Uncertainty. *Marketing Science* 25(1) 25-50.

Hornsey, M.J., J. Jetten. 2004. The individual within the group: Balancing the need to belong with the need to be different. *Personality Soc. Psych. Rev* 8(3) 248–264.

Jeppesen, L.B., L. Fredricksen. 2006. Why do users contribute to firm-hosted user communities? The case of computer controlled music instruments. *Organization Science* 17(1) 45–63.

Jones, E.E., T.S. Pittman. 1982. Toward a general theory of strategic self-presentation. J. Suls, ed. *Psychological Perspectives on the Self* Vol. 1. Lawrence Erlbaum, Hillsdale, NJ, pp. 231–262.

Jones, Q., G. Ravid., S. Rafaeli. 2004. Information Overload and the Message Dynamics of Online Interaction Spaces: A Theoretical Model and Empirical Exploration. *Information Systems Research* 15(2) 194-210.

Katz, E., F. Lazarsfeld. 1995. *Personal Influence*. The Free Press, New York.

Katz, M.L., C. Shapiro. 1985. Network externalities, competition, and compatibility. *Amer. Econom. Rev.* 75(3) 424–440.

Kirmani. A., P. Wright. 1989. Money Talks: Perceived Advertising Expense and Expected Product Quality. *Journal of Consumer Research* 16(3) 344-353.

Kuksov, D. 2007. Brand value in social interaction. *Management Sci.* 53(10) 1634–1644.

Li, X., L.M. Hitt. 2008. Self-selection and information role of online product reviews. *Information Systems Research* 19(4) 456-474.

Liu, Y. 2006. Word-of-mouth for movies: Its dynamics and impact on box office receipts? *J. Marketing* 70(3) 74–89.

Manski, C. 1993. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* 60(3) 531-542.

Mayzlin, D. 2006. Promotional Chat on the Internet. *Marketing Science* 25(2) 157–65.

McPherson, M., L. Smith-Lovin, J.M. Cook. 2001. Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology* 27 415-44.

Moorthy, S., K. Srinivasan. 1995. Signaling quality with a money-back guarantee: the role of transaction costs. *Marketing Science* 14(4) 442–466.

Nair, H., P. Chintagunta, J.-P. Dubé. 2004. Empirical analysis of indirect network effects in the market for personal digital assistants. *Quant. Marketing Econom* 2(1) 23–58.

Narayanan, S., P. Chintagunta, E. Miravete. 2007. The role of self selection, usage uncertainty and learning in the demand for local telephone service. *Quantitative Marketing and Economics* 5(1) 1-34.

Nelson, P.J. 1970. Information and consumer behavior. *Journal of Political Economy* 78(2) 311–329.

Rabe-Hesketh, S., A. Skrondal, A. Pickles. 2004a. Generalized multilevel structural equation modeling. *Psychometrika* 69(2) 167-190.

Rabe-Hesketh, S., A. Skrondal, A. Pickles. 2004b. GLLAMM Manual. U.C. Berkeley Division of Biostatistics Working Paper Series.

Resnick, P., R. Zeckhauser, E. Friedman, K. Kuwabara. 2000. Reputation systems. *Comm. ACM* 43(12) 45–48.

Rimé, B., B. Mesquita, P. Philippot, S. Boca. 1991. Beyond the emotional event: Six studies on the social sharing of emotion. *Cognition and Emotion* 5 435–465.

Rimé, B., P. Philippot, S. Boca, B. Mesquita. 1992. Long-lasting cognitive and social consequences of emotion: Social sharing and rumination. *European Review of Social Psychology* 3 225–257.

Rimé, B., C., Finkenauer, O. Luminet, E. Zech, P. hilippot. 1998. Social sharing of emotion: New evidence and new questions. *European Review of Social Psychology* 9 145-189.

Shapiro, C., H. Varian. 1999. *Information Rules*. Boston: Harvard Business School Press.

Skrondal, A., S. Rabe-Hesketh. 2004. Generalized Latent Variable Modeling: Multilevel, Longitudinal and Structural Equation Models. Boca Raton, FL: Chapman & Hall/ CRC Press.

Sproull, L. 2003. Online Communities. *The Internet Encyclopedia*. H. Bidgoli, ed. John Wiley & Sons, New York.

Sun, B., J. Xie, H.H. Cao. 2004. Product strategy for innovators in markets with network effects. *Marketing Sci.* 23(2) 243–254.

Swann, W.B. 1983. Self-verification: Bringing social reality into harmony with the self. J. Suls, A. G. Greenwald, eds. *Social Psychological Perspectives on the Self*, Vol. 2. Lawrence Erlbaum, Hillsdale, NJ, pp. 33–66.

Tirrell, M. 2009. Netflix, Facebook Link to Show Users' Film Ratings. *Bloomberg*. Available at http://www.bloomberg.com/apps/news?pid=20601103&sid=aCS3VS5zJwvQ&refer=us

Tyler, T.T., P. Degoey, H. Smith. 1996. Understanding why the justice of group procedures matters: A test of the psychological dynamics of the group-value model. *J. Personality Soc. Psych* 70(5) 913–930.

Villas-Boas, J.M. 2004. Consumer Learning, Brand Loyalty, and Competition. *Marketing Science* 23 134-145.

Vives, X. 2008. Informational and Learning in Markets: The impact of market microstructure. Princeton University Press.

Wasserman, S., K. Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge University Press, New York.

Wernerfelt, B. 1995. A rational reconstruction of the compromise effect: Using market data to infer utilities. *J. Consumer Res.* 21(4) 627–633.

Williams, R. 2006. Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata Journal* 6 58–82.

Yang, S., G.M. Allenby. 2003. Modeling Interdependent Consumer Preferences. *Journal of Marketing Research* 40 282–294.

Zhang, J. 2010. The Sound of Silence: Observational Learning in the U.S. Kidney Market. *Marketing Science* 29 315 – 335.

Appendix. GLLAM Estimation Procedure

We follow, mostly, the estimation techniques described in Rabe-Hesketh et al. (2004a) and their Generalized Linear Latent and Mixed Models (GLLAMM) program in STATA (2004b). When latent variables are treated as random and parameters as fixed, the inference is usually based on the marginal likelihood – the likelihood of the data given the latent variables integrated over the latent distribution. The models in GLLAMM include latent or unobserved variables represented by the elements of a vector η and can be interpretable as random effects. In addition, the model is hierarchical to describe the position of a unit of observation. In our case, level-1 is movie units are nested in level-2 units which are online

users.[14] We next explain the GLLAMM estimation procedure in details for Equation (2) in which we assume the random intercept as self-perceived quality, but the same procedure can be applied to other equations with or without the structural equation.

In our model (2), latent true users rating can be written as $R^*_{i,j,t} = v_{i,j,t} + \varepsilon_{i,j,t}$, where $v_{i,j,t} = X'_{i,j,t}\beta + A'_{i,j,t}\delta + \lambda_i \eta^{(2)}_{j,t}$ where the S observed user rating categories a_s, $s = 1,...,S$ are generated by applying thresholds κ_s, $s = 1,...,S-1$ to $R^*_{i,j,t}$ as follows in our case[15]:

$$R_{i,j,t} = \begin{cases} a_1 & if\ R^*_{i,j,t} \leq \kappa_1 \\ a_2 & if\ \kappa_1 < R^*_{i,j,t} \leq \kappa_2 \\ \vdots & \vdots \\ a_S & if\ \kappa_9 < R^*_{i,j,t} \end{cases}$$

where the thresholds κ_s do not vary between movies. If the cumulative density function of $\varepsilon_{i,j,t}$ is F, the cumulative probability τ_s that the response takes on any value up to and including a_s (conditional on the latent and observed explanatory variables) is

$$\tau_s = P(R_{i,j,t} \leq a_s \mid X_{i,j,t}, A_{i,j,t}, \eta^{(2)}_{j,t}) = F(\kappa_s - v_{i,j,t}), \quad s = 1,...,S$$

where $\kappa_1 = -\infty$ and $\kappa_S = \infty$. The probability of the s-th response category is simply then

$$f(R_{i,j,t} = a_s \mid X_{i,j,t}, A_{i,j,t}, \eta^{(2)}_{j,t}) = P(R_{i,j,t} = a_s) = P(\kappa_{s-1} < R^*_{i,j,t} \leq \kappa_s) = F(\kappa_s - v_{i,j,t}) - F(\kappa_{s-1} - v_{i,j,t})$$

We can equivalently write the conditional distribution of user ratings as a cumulative model

$$g(P(R_{i,j,t} < a_s)) = \kappa_s - v_{i,j,t}$$

where $g = F^{-1}$ is the link function.

Then, the model is specified via a family and a link function. If the error $\varepsilon_{i,j,t}$ of the latent rating $R^*_{i,j,t}$ is assumed to have a logistic distribution,

$$Pr(R_{i,j,t} \leq a_s) = Pr(R^*_{i,j,t} \leq \kappa_s) = \frac{\exp(\kappa_s - v_{i,j,t})}{1 + \exp(\kappa_s - v_{i,j,t})}$$

and we have a proportional odds model since the log odds that $R_{i,j,t} \leq a_s$ (conditional on the latent and observed explanatory variables) are

$$\log\left(\frac{Pr(R_{i,j,t} \leq a_s)}{1 - Pr(R_{i,j,t} \leq a_s)}\right) = \kappa_s - v_{i,j,t}$$

---

[14] Third level was time but it was ignored after testing of its significance.

[15] We multiplied observed user rating by 2 to make it be integer in estimation and therefore $a_1 = 1, \ldots a_{10} = 10$.

so that the odds that the response category is less than or equal to a_s for a rating for movie i is a constant multiple of the odds for another rating for the movie i' with odds ratio equal to $\exp(v_{i,j,t} - v_{i',j,t})$ for all s.

The likelihood of the observed data is the likelihood marginal to all latent variables. Let $\boldsymbol{\theta}$ be the vector of all parameters including the regression coefficients β, δ, $\psi^{(2)}$, the threshold parameters κ_s, and the factor loadings λ_i, i=1,…17 , in our model . The number of free parameters in $\boldsymbol{\theta}$ will be reduced if constraints are imposed.

Then, we can specify the likelihood function as

$$\prod_j \int \left\{ \prod_i f(R_{i,j,t} \mid X_{i,j,t}, A_{i,j,t}, \eta_{j,t}^{(2)}; \boldsymbol{\theta}) \right\} g(\eta_{j,t}^{(2)}) d\eta_{j,t}^{(2)}$$

where $g(\eta^{(2)}_{j,t})$ is the prior density of the latent variables. We assume it has a normal distribution with zero mean and variance $\psi^{(2)}$. The GLLAMM program in STATA maximizes the numerically integrated marginal log-likelihood using a Newton-Raphson algorithm.