Self-Selection Bias in Reputation Systems

Mark A. Kramer
MITRE Corporation, 202 Burlington Road, Bedford, MA 01730 USA
mkramer@mitre.org

Abstract. Reputation systems appear to be inherently biased towards better-than-average ratings. We explain this as a consequence of self-selection, where reviewers are drawn disproportionately from the subset of potential consumers favorably predisposed toward the resource. Inflated ratings tend to attract consumers with lower expected value, who have a greater chance of disappointment. Paradoxically, the more accurate the ratings, the greater the degree of self-selection, and the faster the ratings become biased. We derive sufficient conditions under which biased ratings occur. Finally, we outline a potential solution to this problem that involves stating expectations before interaction with the resource, and expressing subsequent ratings in terms of delight or disappointment.

1 Introduction

Trust management involves several different functions: helping a system determine whether to grant a consumer access to a resource ("hard" trust), helping to enforce behavioral norms by providing accountability (sanctioning), and helping a consumer decide whether to employ a resource (signaling). Signaling represents conveyance of information to the consumer about a resource, in support of a decision on whether to employ the resource (which can be practically any service, information, or artifact). The signal must contain information that allows future consumers to estimate the value of the resource, for example, by expressing the likelihood of success of the transaction, the quality of the artifact, or the nature of the information.

Typically, reputation scores are based on reviews of consumer-resource interactions. Because reviews are provided only by the subset of consumers who have selected and interacted with the resource, the group of reviewers may not be representative of the larger group of potential consumers.

Self-selection bias is a classic experimental problem, defined as a false result introduced by having the subjects of an experiment decide for themselves whether or not they will participate [1]. The effect is that the test group may not be representative of the ultimate target population, and therefore the experimental

results cannot be extrapolated to the target population. This is precisely the case in most reputation systems. Reviewers are disproportionately drawn from the subset of potential consumers who are favorably predisposed toward the resource, making it difficult to extrapolate the result to the general population. The self-selection effect in consumer ratings has been previously noted by Li and Hitt [2], but not thoroughly explored.

It is easy to observe positive bias in reputation and rating systems. For example, the average user rating on NetFlix [3] is 3.6 out of 5.0[1]. On Amazon.com, it is 3.9 out of 5.0 [2]. To put the issue in sharper focus, NetFlix users rate SpongeBob SquarePants videos approximately 4 out of 5 stars (Fig. 1). As popular as this cartoon may be among 6-12 year-olds, it is unlikely that the average user of NetFlix would concur with this rating. If the rating seems out of line with expectations, then what value is this rating, and to whom? What "discount" must be applied, if you suspect you are not demographically matched with the average reviewer? Does this rating indicate you might be pleasantly surprised, or severely disappointed?

There might be a tendency to downplay the problem of biased ratings, on the grounds that (a) you already "know" whether or not you would like the SpongeBob movie, (b) you could look at written reviews, or (c) one could get personalized guidance from a recommendation engine. Clearly, if you adore the denizens of Bikini Bottom, then neither reviews nor recommendations are necessary. However, the ubiquity of reviews is evidence that our prior knowledge has limits, and we do not always "know" what we want without them. Surveys of web consumers conducted by BizRate indicate that 44% consult opinion sites before making an online purchase, and 59% consider consumer reviews more valuable than expert reviews [4]. As far as using written reviews instead of ratings, it is true that better choices may result if one can discern the nature of the resource and the motivations or biases of the writer from the review. However, there is every reason to believe that bias pervades opinions expressed in written reviews as much as numerical ratings, and hence we believe the key arguments of this paper apply equally to qualitative and quantitative ratings. In addition, discarding quantitative ratings would eliminate a convenient shorthand and time saver; it may be impractical to read enough reviews to draw appropriate conclusions. Finally, recommendation engines may guide you (as an adult) to more suitable fare than SpongeBob, but even so, reviews and ratings of the recommended movies still play a role in your decisions. No recommendation engine will ever totally replace browsing as a method of finding resources.

In this paper, we explore the reasons that reputation management systems (RMS) are inherently biased, and introduce the *paradox of subjective reputation*, which can be stated as follows: accurate ratings render ratings inaccurate. The nub of the

[1] This number was calculated from over 100 million user ratings collected between October 1998 and December 2005 using the dataset provided for the NetFlix Prize competition. For details, see http://www.netflixprize.com.

paradox is that, while the purpose of a RMS is to support self-selection (allowing consumers to match themselves with resources they value the most); achieving that purpose results in biased reviews, which prevents the RMS from achieving its purpose. The practical effect of this paradox is overly-optimistic ratings driving elevated levels of consumer disappointment.

We begin by creating a model of the self-selection process, and show that under mild assumptions, ratings will be biased. We then explore the dynamics of ratings over time, and present evidence of the effect. Finally, we suggest ways of creating rating systems resistant to self-selection bias.

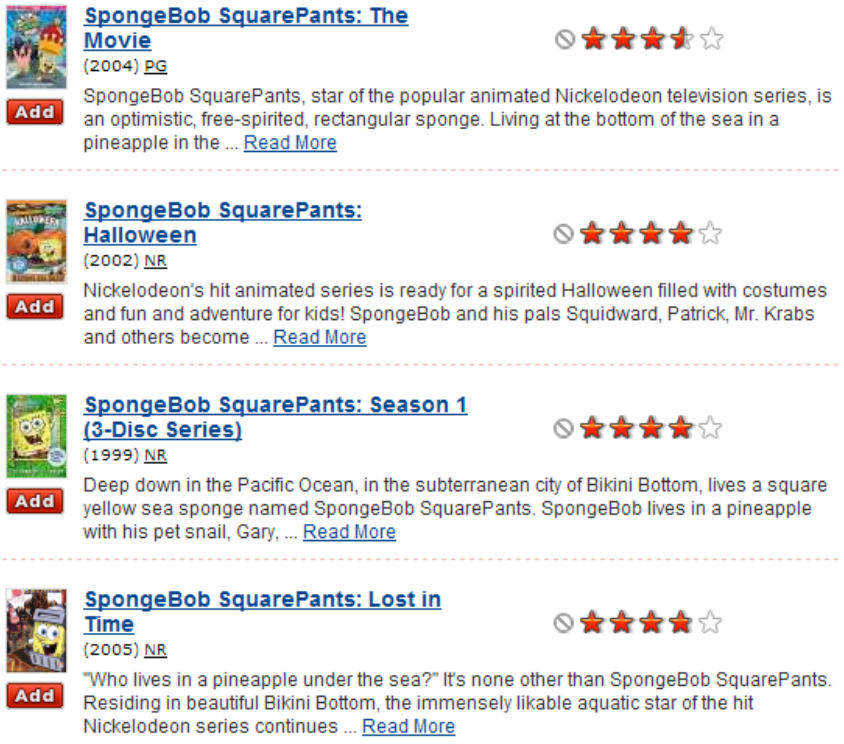

Fig. 1. SpongeBob boasts four-star ratings, but does he deserve it?

2 Expectation and Self-Selection

2.1 Model of Self-Selection

Self-selection happens at a number of different stages in the resource selection process. It occurs when a consumer decides to seek a certain type of resource, when the consumer selects one or more resources for further investigation, when the

consumer selects a specific resource to employ, and finally when the consumer decides to review or provide feedback about the resource. For the purposes of analyzing the phenomenon of self-selection, we are concerned with two populations: the population evaluating a resource (evaluation group $\mathcal{E}$), and the population providing ratings and reviews of the resource (feedback group $\mathcal{F}$). The feedback group might not be a representative sample of those employing the resource; for example, those who are particularly pleased or disappointed might be more likely to provide reviews. However, for simplicity, we will consider population $\mathcal{F}$ to be statistically identical to the population employing the resource.

A typical RMS captures the reviews from the feedback group and provides this data to the evaluation group. As indicated above, $\mathcal{F}$ is not a random sample of $\mathcal{E}$; rather it is a self-selected group containing individuals who, on average, value the resource more highly on average than members of group $\mathcal{E}$. Therefore the ratings awarded by group $\mathcal{F}$ do not represent the latent distribution of opinions in $\mathcal{E}$.

To model this situation, define:

R = resource selected
E = expected satisfaction with the resource
S = actual satisfaction with the resource
P(S) = probability of satisfaction in the evaluation group
$P(S_F) = P(S|R)$ = probability of satisfaction in the feedback group

R, E, and S are propositional variables, either true or false. P(S) represents a hypothetical probability that would result if every member of the evaluation group would employ and rate the resource. P(S) represents the "right" answer for someone in the evaluation mode, in the sense that it represents the likelihood that a consumer will get a satisfactory outcome, independent of the decision whether to employ the resource. Since P(S) is not observable, the question is whether $P(S_F)$ is a reasonable proxy for P(S).

In real life, consumers base their decisions on whether or not to employ a resource on indicators such as price, reputation, and apparent quality, transmitted via advertisement, word-of-mouth, and reviews. This information helps the consumer form a preliminary opinion of the resource, which we represent as the expected satisfaction, E. Because of differences in values, tastes, and priorities, there will be a distribution of expectations within the evaluation group.

If there is a strong expectation of satisfaction, a consumer will be more likely to select the resource. In our binary satisfaction model, we describe self-selection in term of the inequality:

$$P(R|E) > P(R|\sim E) \qquad \text{(Self-selection)}$$

This simply says, in a group of consumers, those that expect to be satisfied with a resource are more likely to select the resource than those who do not expect to be satisfied with the resource. If these expectations turn out to be more right than wrong, consumer expectations will correlate with the achieved satisfaction, S, after employing the resource:

$P(S|E) > P(S|\sim E)$ (Realization of expectations)

As shown in the Appendix, *these two simple conditions are sufficient to prove the resulting feedback will be biased*, *overestimating the satisfaction in the resource*. Bias is defined as the probability of satisfaction in the feedback group being greater than the satisfaction in the evaluation group:

$P(S_F) = P(S|R) > P(S)$ (Biased Rating)

While the proof given in the Appendix shows that bias is a mathematical consequence of the two prior inequalities, the effect can be readily understood without formulae. Consider choosing a movie. The consumers are exposed to some prior information, e.g. a movie review, which appeals to some consumers more than to others. The consumers who expect to like the movie are the ones most likely to see it, and when they see it, they are more likely to enjoy it than those who chose not to see it. In the end, the opinions of the viewers are fed back to the pool of available information. This is illustrated in Fig. 2.

In the following, we quantify the cost of bias in terms of *dissatisfaction* and *lost opportunity*. Dissatisfaction is defined as the probability of not being satisfied after selecting the resource, $P(\sim S|R)$. Lost opportunity is defined as not employing the resource when the consumer would have been satisfied with the resource, $P(S|\sim R)$.

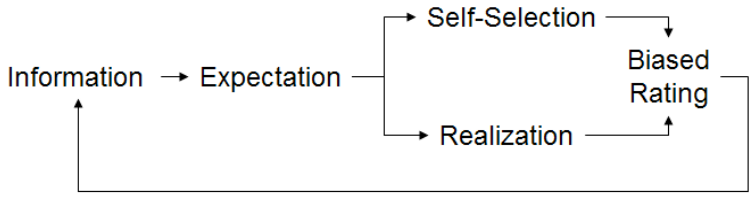

Fig. 2. Causal model of self-selection bias with feedback of ratings

2.2 Effect of Self-Selection on Ratings

If there were no information to form expectations, then consumers could do no better than random selection of resources. If such were the case, the feedback group would be a random sample of the overall population; the reviews would reflect the opinions of the overall population. In this case, reviews would be fair, but disappointment would be maximized, since there would be no opportunity for self-selection. In the other extreme, if there were perfect information, consumers would always know in advance if they would be satisfied with the resource, and self-selection would be perfect, reviews would be uniformly glowing, and there would be no dissatisfaction whatever.

Anchored by these two extremes, it can be seen that increasing (accurate) information increases self-selection, increases ratings bias, and decreases dissatisfaction. Conversely, reduced, biased, or inaccurate information decreases

self-selection, decreases ratings bias leading to fairer ratings, and increases dissatisfaction.

To put this in perspective, imagine what could happen when 10 people land on the NetFlix page shown in Fig. 1:

- Three consumers are SpongeBob fans who see the movie, and rate it five stars.
- Six consumers don't like SpongeBob, ignore the high ratings, and do not go see the movie.
- One consumer who has no prior opinion about porous invertebrates inhabiting undersea pineapples, is impressed by the high ratings, and sees the movie. He rates the movie one star.

The average new rating is (5+5+5+1)/4 = 4 stars, so the rating remains unchanged; the "trap" remains baited for subsequent consumers. Seven of the ten consumers have reason to be skeptical of the rating system and are less likely to believe it in the future. Nine of ten consumers with strong prior beliefs and depended very little on the ratings system. Only one consumer depended on the ratings system, and to him it was the cause of disappointment, wasted time and money.

If the ratings were unbiased, they would indicate approval by only 3 out of 10 consumers. This data could potentially change the decision of the 10^{th} consumer - or at least, reduce his level of surprise if the movie disappoints.

Example:

A population consists of 100 individuals evaluating a resource. Assume they have enough information to evaluate the resource with 80% accuracy, for both type I and type II errors ($P(S|E) = P(\sim S|\sim E) = 0.8$). Suppose that when these individuals are provided with *unbiased* information about the resource, 50 expect to be satisfied with the resource. For simplicity, assume the same individuals go on to employ the resource. Of the 50 employing the resource, 40 of these individuals will be satisfied. Of the 50 who are not expecting to be satisfied, 10 would have been satisfied if they had elected to employ the resource. With biased information, assume an additional 10 individuals are persuaded to employ the resource. In the feedback group of 60 individuals, 40 of the first 50 are satisfied (as before), but only 2 of the additional 10 are satisfied. Therefore, the probability of satisfaction falls to 42/60, or 70%. Among the remaining 40 consumers not selecting the resource, the lost opportunity is 8/40, or 20%. This is summarized in Table 1.

We see from this Example that biased feedback increases the rate and quantity of disappointed individuals. This is not surprising since biased information decreases the efficiency of self-selection. What is surprising is that the group provided with unbiased information actually produces ratings that are *more* biased than the group presented with biased information (80% positive versus 70% positive). This is because unbiased (accurate) rating information creates efficient self-selection, which enhances the ratings bias.

Table 1. Data for Example

	Unbiased Information	Biased Information
Evaluating Population	100	100
# Expecting satisfaction (E)	50	60
# Selecting resource (R)	50	60
Feedback group satisfaction	40/50 = 80%	42/60 = 70%
Disappointment	10/50 = 20%	18/60 = 30%
Lost opportunity	10/50 = 20%	8/40 = 20%

2.3 Effect of Bias on Self-Selection

In the preceding section, we examined how self-selection affects ratings. In this section, we examine how ratings affect self-selection. Our assumption is that the primary action of biased feedback is to increase the number of consumers employing the resource. Chevalier and Mayzlin [5] have shown that online book ratings do affect book sales. The consumers most likely to be influenced by biased feedback are those without strong preexisting opinions. As a group, these "swing" consumers have lower expectations than the group who would select the resource based on unbiased feedback. If expectations are well-calibrated, the likelihood of dissatisfaction in the "swing" group will be higher than in the first feedback group. By delving deeper into the group of consumers, bias tends to decrease the selectivity of the feedback group. This is consistent with the previous observation that less (or inaccurate) information decreases self-selection, and results in less biased ratings.

As shown in Fig. 2, ratings systems involve a feedback loop. It is a negative feedback loop because increasing information tends to increase self-selection, which tends to increase ratings bias, which tends to decrease information. Systems with negative feedback can show a variety of interesting dynamics, including overdamped (asymptotic approach to steady state) and underdamped responses (overshoot followed by asymptotic approach to steady state).

To demonstrate the effect of feeding back biased ratings, we have to use a more complex model than the binary satisfaction model used above. Assume the following simple deterministic situation:

- A resource whose latent satisfaction (S) is uniformly distributed between 0 and 100
- Perfectly-calibrated consumer expectations (E=S)
- Average rating equal to average satisfaction
- Sequential subgroups of 100 consumers
- Number selecting the resource in each subgroup proportional to the average rating thus far received, i.e. if the resource has earned a perfect rating of 100, then all consumers in the subgroup will select the resource
- Initial group of 10 random "pioneers" rating the resource

In this situation, we might expect an average rating of 50, since this corresponds to the average latent satisfaction of all consumers. Furthermore, the initial rating of the resource is fair (50), because the pioneers are randomly selected. In the round immediately following the pioneers, 50 consumers whose expectation exceeds 50 employ the resource. Among this group, the average rating is 75. Thus, the cumulative average rating rises to (50*10 + 75*50)/60 = 70.8. This demonstrates the paradox: *accurate ratings render ratings inaccurate*. Table 2 shows the evolution of the average rating through five rounds of consumers, and shows that the steady state is reached at cumulative average rating of 67.

A variation on this scenario is when the initial group consists of a group of enthusiasts, fans, or shills who award maximum ratings, either as a sincere endorsement or calculated attempt to expand the audience for a resource. In this case, the initial ratings are maximal, which draws large group of consumers in round 2. However, the average rating plummets when many in the group are disappointed (Table 3).

Table 2. Dynamic evolution of ratings seeded by random pioneer group

Round	Total Subgroup Size	# Selecting Resource	Average Rating	Cumulative Average Rating
1 (Pioneer Group)	10	10	50	50
2	100	50	75	70.8
3	100	70	65	67.7
4	100	67	66.5	67.3
5	100	67	66.5	67.1
6	100	67	66.5	67.0
Steady state	100	66	67.0	67.0

Table 3. Dynamic evolution of ratings seeded by shill (or enthusiast) group

Round	Total Subgroup Size	# Selecting Resource	Average Rating	Cumulative Average Rating
1 (Shill Group)	10	10	100	100
2	100	100	50	54.5
3	100	54	73	60.6
4	100	60	70	63.1
5	100	63	68.5	65.0
6	100	64	68.0	65.4
Steady state	100	66	67.0	67.0

Figure 3 shows this data in graphical form, for two initial conditions (pioneer and shill), and different subgroup sizes. The larger the subgroup, the larger the overshoot effect in the opposite direction from the initial rating. In both cases, the final steady state is approximately 67/100 (slight differences are due to round-off effects). We

can see this analytically, because given a fraction f of the subgroup, the average rating is given by r = 1 - f/2. If the rating draws an equivalent fraction of the subgroup, then f = r, so r = 1 - r/2, or r = 0.67. Also notes that small subgroups lead to overdamped behavior, while larger subgroups lead to underdamped (overshoot) behavior.

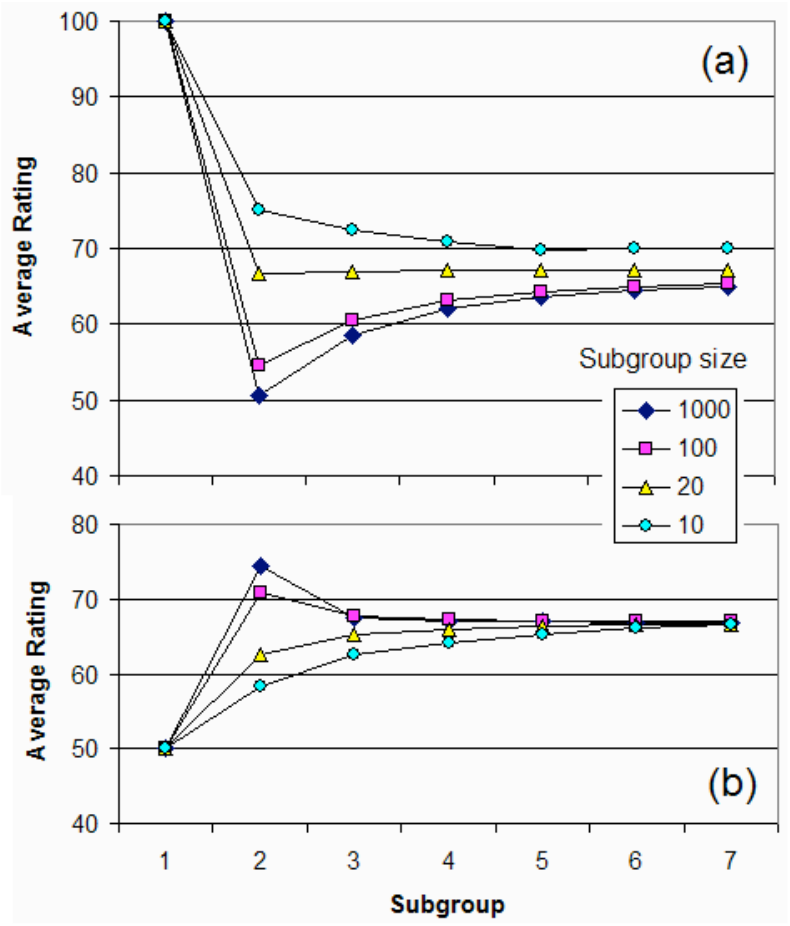

Fig. 3. Dynamic evolution of ratings as a function of subgroup size for (a) initial shill (or enthusiast) group and (b) initial pioneer (fair rating) group

2.4 Steady-State Bias

Steady state is achieved at the point where the available information, including ratings, recruits a new group of consumers whose composition is such that the average rating from the new group matches the existing rating. The steady state is the fixed point of the function r = R(G(r)), where g = G(r) is a function that generates

a group of consumers employing the resource given an average rating r, and r = R(g) is a function that generates ratings for a group, g. As we have already shown, under a few easily-satisfied assumptions, this fixed point is biased in favor of the resource. How large is the steady-state bias?

Suppose the distribution of user expectations is given by a standard normal distribution, and the final ratings are correlated to the expectations via a correlation coefficient between 0 and 1. Assume a fraction f of the overall population, drawn from the top of the expectation distribution, become the reviewers. In this case, we can determine the bias between the unbiased rating and the observed rating through simulation, where we generate a Gaussian distribution of expectations, select the reviewers from the top of the distribution, and simulate their final ratings according to the given correlation between expected and actual ratings.

Figure 4 shows the results of this simulation. Bias is higher when a smaller fraction of the population selects the resource, and higher with stronger correlation between expected and actual ratings. Without realization of expectations, there is no bias.

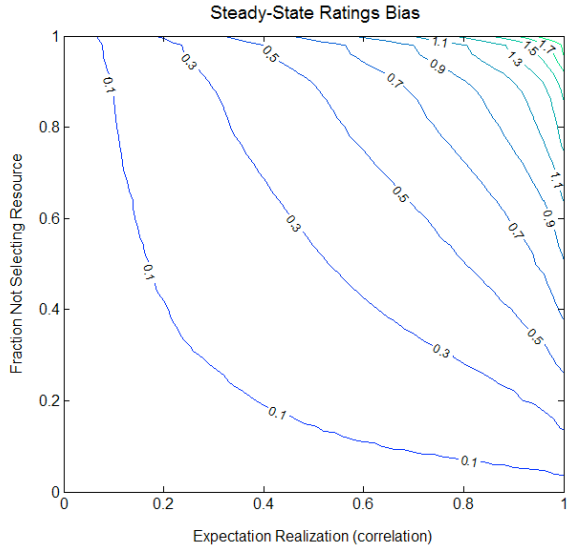

Fig. 4. Extent of steady-state ratings bias for normal distribution of expectations, as a function of correlation between expected and actual ratings (x-axis), and the fraction of population not selecting the resource (y-axis).

3. Evidence from Ratings Systems

At this time, the existence of bias due to self-selection can be considered a hypothesis. However, the predictions are easily testable, by comparing the ratings

from a group of randomly-selected consumers against ratings from a group of self-selected consumers. It is possible that effects ignored here might cancel out the expected bias; for example, fans of SpongeBob might be hypercritical and give lower ratings if the SpongeBob movie is not up to snuff, resulting in ratings equal to or lower than randomly-selected people.

However, if self-selection bias did not exist, one would expect the average rating in rating systems to be close to the median value (2.5/5.0 stars). As mentioned in the introduction, in NetFlix, the average rating is 3.6/5.0. At Amazon, the average book rating is even higher, 3.9/5.0 [2]. People are evidently very satisfied with the books and movies they choose; self-selection is indeed working to some extent.

One can also look at the dynamic evolution of ratings in these systems. Li and Hitt [3] gathered data from 2651 hardback books published from 2000-2004 reviewed on Amazon.com in a 5-month period in 2004. They correlated average rating against the time the since the book was published, correcting for the average rating of a book, and discovered a clear declining trend. The average rating conformed to a negative exponential:

$$\text{Rating for book } i = 3.90 + 0.45 * \exp(-0.746\, t) + \alpha_i$$

where t is the amount of time (in months) after publication, and α_i is the steady-state rating of book i above or below the overall average of 3.9 (out of 5.0). The average rating drop is approximately half a point on this scale. Li and Hitt also conclude that the time-variant component of the rating has a "significant impact on book sales, which leads to the conclusion that consumers did not fully account for the positive bias of early reviewers". If our analysis is correct, the positive bias is not just an early effect, but a steady-state effect as well.

We analyzed data from the NetFlix challenge problem in a similar manner. This data shows an average increase of about 0.2 points (out of 5.0) during the lifetime of a typical movie. A large majority, 765 of 1000 randomly-selected movies, showed an increase in ratings over time. In terms of our model, this suggests that the initial audience is more random than the audience that develops over time -- i.e., it takes time for a movie to "find its audience". It is possible that shill reviews are more common and influential in the book domain than the movie domain.

4. Avoiding Bias in Reputation Management Systems

Since ratings systems have a built-in bias in favor of the resource, alternative designs that are more resistant to self-selection bias are of interest. Personalization is well-known approach improving ratings. The most obvious way to achieve personalization is using demographics, for example, correlating SpongeBob preference to viewer age. However, dividing consumers into demographic subgroups does not eliminate self-selection bias, because within each demographic, self-selection is still the prime determiner of who selects the resource and becomes a reviewer. Furthermore, available demographics might not create useful subsets of consumers with different preferences for a resource (for example, determining who

is interested in a particular technical topic). The other common approach to personalization is collaborative filtering. However, as we argued in the introduction, while consumers may appreciate personalized recommendations, they also expect to be able to discover resources by browsing, consulting both aggregate ratings and individual reviews. The problem of consumers failing to discount biased aggregate ratings (as well as biased written review), does not go away.

In closing, we mention a novel approach for eliminating bias. It involves dividing the reviewers into subgroups according to their prior expectations. Instead of rating the resource in absolute terms, the rating is collected in two parts: the prior expectation E and the posterior satisfaction S. The latter can be collected in terms of surprise (whether the encounter was worse, better, or the same as expected). Collecting these two pieces of data allows the reputation system to build up approximations to the conditional probability P(S|E,R). We have already argued that S is conditionally independent of resource selection (R) given E, and therefore $P(S|E,R) \approx P(S|E)$. Conditioning on E takes the resource selection decision literally and figuratively out of the equation. Making the expectation explicit bridges the gap between the satisfaction of the evaluation group $\mathcal{E}$ and the feedback group $\mathcal{F}$.

Here is one way this approach might work in the context of a movie recommendation system. Consumers browse or use recommendation engines to find and select resources in the typical manner. However, instead of the aggregate rating, data is presented in conditional form:

Among people who thought they would love this movie:
- *40% loved it*
- *30% liked it*
- *20% neither liked nor disliked it*
- *10% disliked it*

Among people who thought they would like this movie:
- *5% loved it…*

When a resource is selected (for example, when a user adds a movie onto his or her queue in NetFlix), he or she is solicited for an expectation. The expectation scale could be the similar to the five-star rating scheme, or a verbal scale ("I think I'll love this movie", "I think I'll like this movie", "I somewhat doubt I'll like this movie", etc.). The elicitation of expectation information can take other forms, for example, asking the viewer if he or she is an "avid SpongeBob fan", "neutral to SpongeBob", etc. or even "dying to see this movie", "looking forward to seeing this movie", or "not looking forward to seeing this movie".

After viewing the movie, feedback can be collected in conventional form, or in terms of delight or disappointment, for example:

I liked this movie:
- *Much more than expected*
- *A little more than expected*
- *About the same as expected*
- *A little less than expected*
- *Much less than expected*

This approach reduces or eliminates self-selection bias because, although the majority of responses are collected from those who expect to like or love the movie, these responses are never pooled with the smaller number of respondents who have lower prior expectations. Therefore, the information represented by these viewpoints is not overwhelmed by sheer numbers.

5. Conclusions

The problem of ratings bias and the market inefficiency (consumer disappointment) that results has not been widely recognized or analyzed. We have shown that if prior expectations exist and are used to select resources, and these expectations positively correlate with results obtained, then biased ratings will result. We have also explored the dynamics of ratings under the assumption that higher ratings attract more consumers. The analysis reveals a paradoxical situation, where biased ratings tend to attract a broader cross-section of consumers and drive the ratings to become less biased, and unbiased ratings tend to attract a focused set of consumers who value the resource highly, which drives towards more biased ratings. These countervailing forces explain the time trends in ratings.

Creating a fair and unbiased rating system remains an open problem. The framework presented here suggests an approach centered on collecting prior expectations, as well as after-the-fact ratings. There is also scope for further investigation into data collected by existing systems to try and determine the extent of actual bias, and to what extent consumers are recruited by biased ratings.

Acknowledgement
The author gratefully acknowledges the sponsorship of MITRE Corporation, thoughtful input from the SezHoo research group (Roger Costello, Andrew Gregorowicz, Dino Konstantopoulos, and Marcia Lazo), and the support of Harry Sleeper, Tom Gannon and Ed Palo.

References

1. J.J. Heckerman, Sample Selection Bias as a Specification Error, *Econometrica*, **47**(1), 153-162 (1997).

2. Xinxin Li and L.M. Hitt, Self Selection and Information Role of Online Product Reviews, Working Paper, Wharton School of Management, University of Pennsylvania (2004); http://opim-sun.wharton.upenn.edu/wise2004/sat321.pdf.

3. Netflix, Inc. (January 17, 2007); http://www.netflix.com.

4. C. Piller, Everyone Is A Critic in Cyberspace, Los Angeles Times (December 3, 1999).

5. J. Chevalier and D. Mayzlin, The Effect of Word of Mouth on Sales: Online Book Reviews, Working Paper, Yale School of Management (2003).

Appendix

As described in the text, we assume consumer expectation predicts selection of the resource, and likewise, expectation predicts satisfaction with the resource:

1) $P(R|E) > P(R|\sim E)$ (self-selection)
2) $P(S|E) > P(S|\sim E)$ (fulfillment of expectations)

From (1), noting that $P(R) = P(R|E)P(E)+P(R|\sim E)P(\sim E) < P(R|E)P(E)+ P(R|E)P(\sim E)$, it follows that $P(R|E) > P(R)$. By Bayes theorem, $P(E|R)P(R)/P(E) > P(R)$, and therefore:

3) $P(E|R) > P(E)$. Combining (2) and (3),
4) $(P(E|R) - P(E))(P(S|E) - P(S|\sim E)) > 0$

Expanding algebraically, and simplifying:

5) $(P(S|E)P(E|R)+P(S|\sim E)-P(S|\sim E)P(E|R)) - (P(S|E)P(E)+P(S|\sim E)-P(S|\sim E)P(E)) > 0$

Noting that $1-P(E|R) = P(\sim E|R)$ and $1-P(E)=P(\sim E)$, then:

6) $(P(S|E)P(E|R)+P(S|\sim E)P(\sim E|R)) - (P(S|E)P(E)+P(S|\sim E)P(\sim E)) > 0$

We can identify the second term as $P(S)$. If we assume that S is conditionally independent of R given E, i.e. $P(S|E,R) = P(S|E)$, the first term is recognized as $P(S|R)$. Conditional independence is a good assumption since once the consumer decides whether he is likely to be satisfied by the resource, the selection decision does not influence the likelihood of being actually satisfied with the resource. Therefore:

7) $P(S|R) - P(S) > 0$, and finally
8) $P(S_F) > P(S)$

This shows that biased feedback (8) will result whenever there is self-selection based on expectations (1) and greater-than-random fulfillment of expectations (2).