Public discourse in the web does not exhibit group polarization

Fang Wu and Bernardo A. Huberman
HP Laboratories
Palo Alto, CA 94304

August 7, 2008

Abstract

We performed a massive study of the dynamics of group deliberation among several websites containing millions of opinions on topics ranging from books to media. Contrary to the common phenomenon of group polarization observed offline, we measured a strong tendency towards moderate views in the course of time. This phenomenon possibly operates through a self-selection bias whereby previous comments and ratings elicit contrarian views that soften the previous opinions.

No aspect of the massive participation in content creation that the web enables is more evident than in the countless number of opinions, news and product reviews that are constantly posted on the Internet. Since these opinions play such an important role in trust building and the creation of consensus about many issues and products, there have been a number of recent of studies focused on the design, evaluation and utilization of online opinion systems [1, 2, 3, 4] (for a survey, see [5]). Given the importance of group opinions to collective social processes such as DELPHI, group polarization and information cascades [6, 7, 8, 9, 10] it is surprising that with the exception of one study [11], little research has been done on the dynamic aspects of online opinion formation. It remains unclear, for example, whether the opinions about books, movies or societal views fluctuate a long time before reaching a final consensus, or they undergo any systematic changes as time goes on. Thus the need to understand how online opinions are created and evolve in time in order to draw accurate conclusions from that data.

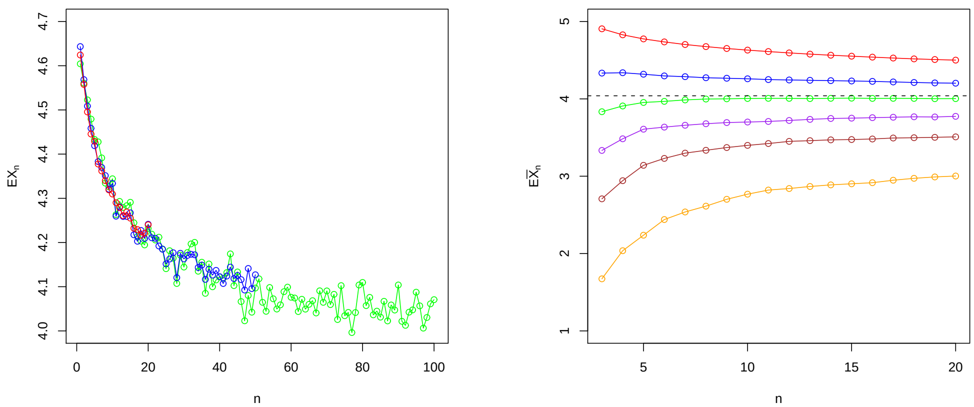
It is well known that in the case of group polarization, members of a discussion group tend to move and coalesce not toward the middle of antecedent dispositions, but toward a more extreme position in the direction indicated by those dispositions [12, 13, 14]. Within this context we studied the dynamics of online opinion expression by analyzing the temporal evolution of a very large set of user views, ranging from online reviews of the best selling books at `Amazon.com`, and movie reviews at the Internet Movie Database `IMDB`. Surprisingly, our analysis of online opinions revealed a trend that runs counter to the group polarization phenomenon. On the massive scale that the web offers, we observed that later opinions in the course of time tend to show a large difference with previous ones which moderates the average opinion to the less extreme. This is a robust and quantitative observation for which we can only offer a tentative explanation in terms of the cost of expressing an opinion to the group at large.

In order to perform this study we first analyzed book ratings posted on `Amazon.com`. Our sample consisted of 1,729,830 book ratings of the top 4,000 best-selling titles of `Amazon` in each of the following 12 categories, as of July 1, 2007: arts & photography, biographies & memoirs, history, literature & fiction, mystery & thrillers, reference, religion & spirituality, sports, travel, nonfiction, science, and entertainment. For each of the 48,000 books, a series of user ratings was collected in time order, where each rating is an integer between 1 and 5. Among the 48,000 books, 16,454 books have no less than 20 ratings, 7,385 books have no less than 50 ratings, and 3,495 books have no less than 100 ratings. We first checked the average rating of books (EX_n) in each of these three sets of books as a function of the rating index (n).[1] As can be seen from Fig. 1(a), the ratings are generally favorable (the average of all 1,729,830 ratings is $\bar{X}_{\text{all}} = 4.04$). There is a clear decreasing trend, indicating that later users tend to write different reviews from those of earlier users. This is consistent with the empirical finding of [15] that when group members voice their opinions sequentially, an opinion shift may be induced. However, as opposed to the widely-observed group polarization effect, the overall opinion on `Amazon` tends to *decrease* away from the extreme ones.

Since the book ratings in our dataset are generally favorable, one might wonder whether the moderating effect in Fig. 1(a) is merely an artifact of high initial ratings. It would be helpful if one could locate a set of books with low initial ratings, and see if their ratings also exhibit a moderating effect over time. To this end we divided the books into 6 tiers based on the average value of their first 3 ratings ($\bar{X}_3$), and plotted the average rating as a function of the rating index for each tier. As one can see in Fig. 1(b), all 6 tiers tend to move closer toward the overall mean $\bar{X}_{\text{all}} = 4.04$, as opposed again to the divergent pattern commonly observed in group polarization experiments.

In order to better compare our results with existing studies of group polarization which usually measure the difference between the final and initial average opinion, we next measured the change of ratings at the level of each individual book. For each book with $l \geq 2n$ ratings, we calculated the mean of its first n ratings ($\bar{X}_{1,n} = \frac{1}{n}\sum_{i=1}^{n} X_i$) and the mean of its last n ratings ($\bar{X}_{l-n+1,l} = \frac{1}{n}\sum_{i=l-n+1}^{l} X_i$). We then performed a series of statistical tests on the alternative hypothesis "$\bar{X}_{l-n+1,l} < \bar{X}_{1,n}$" for books in each tier, n ranging from 1 to 50. The results are listed in Table 1 (see also Fig. 2). As can be

[1] We will use the following notations throughout this paper: X_n denotes the quantified value of the n'th opinion of one particular item; $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ denotes the average value of the first n opinions of one particular item; EX_n denotes the sample mean of X_n for all items in a dataset; $E\bar{X}_n$ denotes the sample mean of $\bar{X}_n$ for all items in a dataset.



(a) (b)

Figure 1: (a) Average rating of `Amazon` book reviews as a function of the rating index. EX_n is the sample average rating of all the n'th ratings of books with no less than (1) 20 ratings (red), (2) 50 ratings (blue), and (3) 100 ratings (green). As one can see from the figure, EX_n decreases by 0.4 stars in 20 steps. (b) We divided the 16,454 books which have no less than 20 ratings into 6 tiers. From top to bottom respectively, they contain 11,393 books with $\bar{X}_3 \in (4.5, 5]$, 1,645 books with $\bar{X}_3 \in (4, 4.5]$, 2,252 books with $\bar{X}_3 \in (3.5, 4]$, 500 books with $\bar{X}_3 \in (3, 3.5]$, 562 books with $\bar{X}_3 \in (2, 3]$, and 102 books with $\bar{X}_3 \in [1, 2]$. The horizontal dashed line $X_{\text{all}} = 4.04$ is the overall mean of all 1,729,830 ratings in our full dataset. We see that the average book rating in each tier shifts towards the dashed line.

Tier	$\bar{X}_3$	Result
1	$(4.5, 5]$	Rejects null with $p < 0.001$ for $n = 1, \ldots, 50$
2	$(4, 4.5]$	Rejects null with $p < 0.001$ for $n = 1, \ldots, 50$
3	$(3.5, 4]$	Insignificant
4	$(3, 3.5]$	Supports null with $p < 0.05$ for $n = 1, \ldots, 17$
5	$(2, 3]$	Supports null with $p < 0.05$ for $n = 1, \ldots, 33$
6	$(1, 2]$	Supports null with $p < 0.05$ for $n = 1, \ldots, 50$

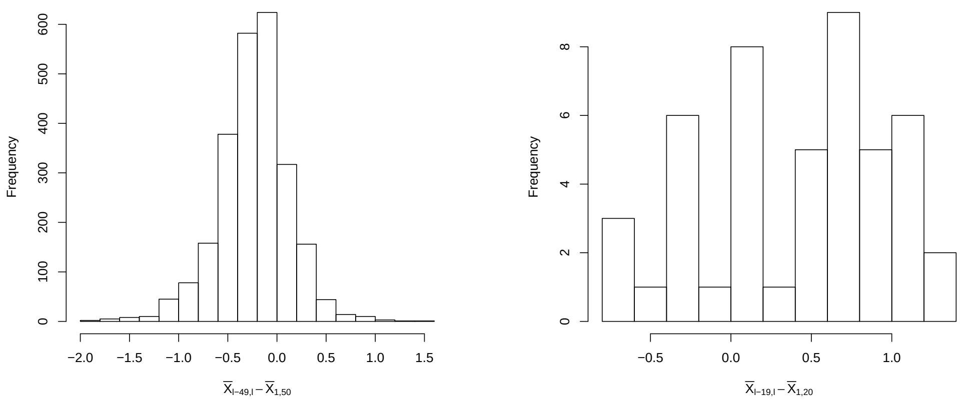
Table 1: Test of the null hypothesis $\bar{X}_{l-n+1,l} > \bar{X}_{1,n}$ for books in 6 tiers.

seen, the conclusions drawn from Fig. 1(b) are robust.

While these empirical findings reveal interesting trends in online opinion formation in contrast to group polarization, a more careful examination will detect that the two experimental settings—ours versus group polarization experiments—are not exactly the same. First of all, the subjects in a group polarization experiment all have to engage in a discussion, but in our study the participation level of each user is unknown. (For example, we do not know whether an early reviewer comes back to read later reviews.) Second, online discourse tends to be anonymous and not face-to-face.[2] Third, in group polarization experiments consensus is often reached at the end of the discourse,[3] but on `Amazon` there is no such requirement. In particular, the convergence pattern shown in Fig. 1(b) does not necessarily imply that people come to an agreement in the end. In fact, when we measured the standard deviation $\sigma_{1,n}$ of the first n ratings and $\sigma_{l-n+1,l}$ of the last n ratings for all books with no less than $2n$ ratings, all 49 tests supported the alternative hypothesis "$\sigma_{l-n+1,l} > \sigma_{1,n}$" with p-value less than 0.001 for $n = 2, \ldots, 50$, suggesting diversification of opinions over time. We also measured the fraction of each star rating as a function of the rating index. Let $f_n(r)$ denote the fraction of r-star ratings among all the n'th ratings in our dataset, for $r = 1, 2, 3, 4, 5$. Fig. 3 plots the growth of fractions for $r = 1, 2, 3, 4$. As one

[2]Recent studies of computer-mediated communication have shown that anonymity and de-individuation usually enhances group polarization [16, 17]. This is in contrast with our findings.

[3]This is not required. In cases where consensus is not reached, the final group decision is usually measured by the average position of all group members' final position [18, 19].



(a) (b)

Figure 2: Sample histograms of statistical tests listed in Table 1. (a) Histogram of $\bar{X}_{l-49,l} - \bar{X}_{1,50}$ for books in the top tier. Notice the shift towards the left of the origin, signaling a downward trend in the average opinion. (b) Histogram of $\bar{X}_{l-19,l} - \bar{X}_{1,20}$ for books in the bottom tier. Notice that the majority of the data is to the right of the origin, indicating an upward trend in the average opinion.

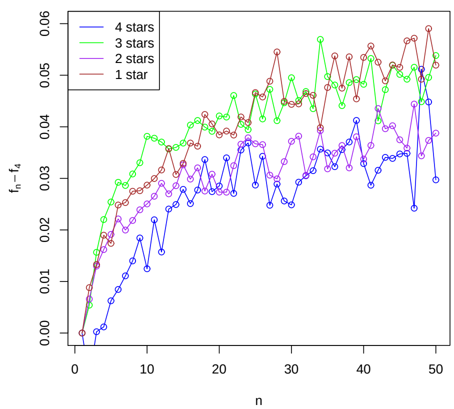
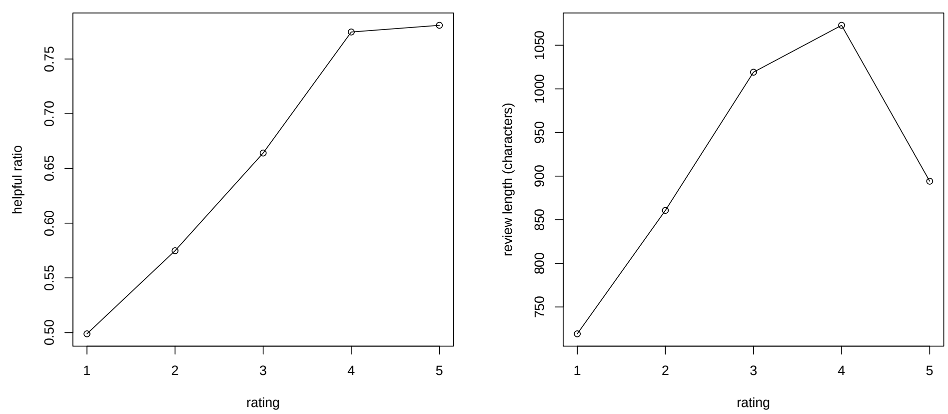


Figure 3: The change in the fraction of 4-star, 3-star, 2-star, and 1-star reviews on `Amazon`.

can see the fraction of 1-star ratings grows fastest, indicating that instead of ending with a group of people with moderate opinions, one observes a variety of diverse views. This shows that public discourse on `Amazon` typically intensifies over time and does not reach a consensus at the end, in contrast to the imitative behavior seen in herding and information cascades.

Perhaps the most salient difference between our study and that of group polarization is that the degree of group polarization is traditionally measured by the difference between the average final position and the average initial position of the *same* group of subjects, whereas in Fig. 1(b) we could only compare the ratings of *two* groups of subjects (the first n and the last n), for `Amazon` does not allow its users to leave multiple reviews. Hence our results do not necessarily imply that as time goes on the average opinion of the whole population changes, for the late reviewers might come from a different group than the earlier ones and need not be representative of the whole population. This is seen when plotting the average "helpful ratio" as a function of star rating in Fig. 4 for users of `Amazon`. As can be seen, the whole population finds high ratings in general more helpful than low ratings. If one assumes that agreeing with a review is correlated with finding it useful, this implies that the majority of the population does not necessarily agree



(a) (b)

Figure 4: (a) The average helpful ratio of five different star ratings. (b) The average review length of five different star ratings in the number of characters. The data is calculated for 4,000 bestselling mystery books. By comparing the two figures it is clear that people find high ratings more helpful not just because they are longer. For instance, 5-star reviews are on average shorter than 4-star and 3-star reviews but are nevertheless more helpful.

with the low ratings. This additional data then suggests that rather than indicating a real choice shift in the whole population, the observed dynamic trend on `Amazon` might be more of a *selection bias*. By this we mean that the more recent ratings are generated by a subgroup of users whose opinions deviate considerably from the average opinion of the whole group, a point that we will elaborate below.

On reflection, it is rather surprising that people contribute opinions and reviews of topics which have already been extensively covered by others. While posting views is easy to understand when it involves little effort, like clicking on a button of a website, it is more puzzling in situations where it is costly, such as composing a review.[4] If the opportunity to affect the overall

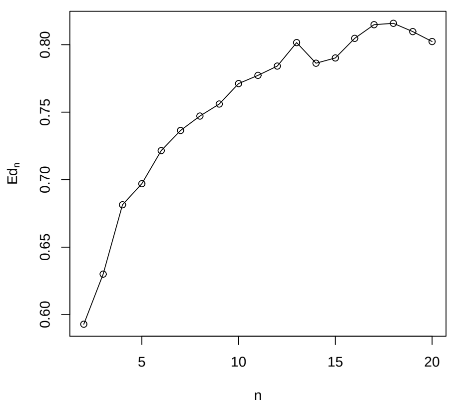
[4]When a user of `Amazon` decides to review a book, she is required to write a short paragraph of review in addition to a simple star rating. The average word count of `Amazon` reviews is 181.5 words [22], so the cost of opinion expression is indeed high.

opinion or rating diminishes with the number of published ones, why does anyone bother to incur the cost of contributing yet another review? From a rational choice theory point of view, if the utility to be gained does not outweigh the cost, people would refrain from expressing their views. And yet they do not. This is reminiscent of the well analyzed voter's paradox [20, 21], where a rational calculation of their success probability at determining the outcome of an election would make people stay home rather than vote, and yet they show up at the polls with high turnout rates. In contrast to a political election, there is no concept of winning in online opinion systems. Rather, by contributing her own opinion to an existing opinion pool, a person affects the average or the distribution of opinions by a marginal amount that diminishes with the size of that pool.

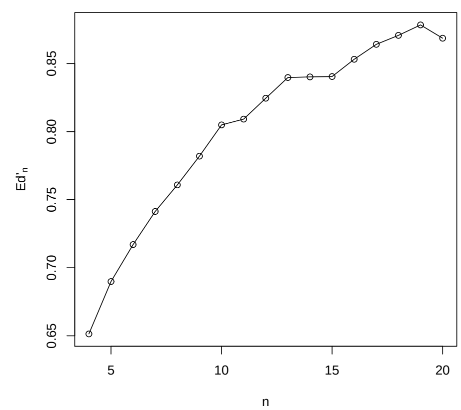
One possible explanation for these results is that in cases like `Amazon`, people will derive more utility the more they can influence the overall rating, as in the voter's paradox. To be precise, in cases where users' opinions can be quantified and aggregated into an average value (when a user opens up a book page on `Amazon`, she sees the average rating at the top of the page right under the title), the influence of an online opinion can be measured by how much its expression will change the average opinion. Suppose that n users have expressed their opinions, $X_1, \ldots, X_n$, on a given topic at a website. If the $(n+1)$'th person expresses a new opinion X_{n+1}, it will move the average rating to $\bar{X}_{n+1} = (n\bar{X}_n + X_{n+1})/(n+1)$, and the absolute change in the average rating is given by $|\bar{X}_{n+1} - \bar{X}_n| = |X_{n+1} - \bar{X}_n|/(n+1)$. Thus a person is more likely to express her opinion whenever $|X_{n+1} - \bar{X}_n|$ is large—an opinion is likely to be expressed if it deviates by a significant amount from those already stated. Indeed, what is the point of leaving another 5-star review after one hundred people have already done so?[5]

In order to test this hypothesis, we measured directly how much one's rating deviates from the observed average rating. We plot the expected

[5]This point has also been made within the "brag-and-moan" model [4, 23] which assumes that consumers only choose to write reviews when they are very satisfied with the products they purchased (brag), or very disgruntled (moan). Note however, that the brag-and-moan model is static and thus predicts that $\bar{X}_n$ is constant over time, in contradiction with the observed dynamical trends.



(a)



(b)

Figure 5: (a) The average deviation of `Amazon` ratings increases with the number of reviewers. (b) The average deviation of X_n from $(X_{n-3} + X_{n-2} + X_{n-1})/3$ increases with the number of reviewers.

deviation $Ed_n = E|X_n - \bar{X}_{n-1}|$ as a function of n in Fig. 5(a), where X_n is the rating left by the n'th user, and $\bar{X}_{n-1}$ is the average rating the n'th user observes. As can be seen, Ed_n increases with n. Since the expected deviation Ed_n of an i.i.d. sequence normally *decreases* with n, this increasing trend is indeed significant. Because `Amazon` also shows its users a histogram of all the past ratings and a list of the most recent ratings near the bottom of the page, we also measured the deviation of each rating from its recent three predecessors.[6] Fig. 5(b) plots $Ed'_n = E|X_n - (X_{n-3} + X_{n-2} + X_{n-1})/3|$ as a function of n. Not only the increasing trend remains but the deviation actually exceeds Ed_n by 0.05 stars. This again supports our conjecture that those users who disagree from the existing average will be more willing to express themselves and thus soften the overall opinion of a given book.

While our hypothesis seems to explain the softening of opinions observed

[6]It is known in the group polarization literature that exposure to the group average is sufficient to stimulate a more polarized response, and that those exposed to the full distribution of others' choice are not significantly more polarized than those who merely witnessed the group average [24, 25].

in Amazon, it would be more conclusive if one could conduct a test that directly compares people's opinions expressed at different cost levels. In order to address this issue we conducted a study of IMDB.com (The Internet Movie Database). Unlike users of Amazon who are required to write a review when rating a book, users of IMDB are free to *choose* the effort level when reviewing a movie. Specifically, after observing the current average rating of a movie, a user can either submit a quick rating by clicking on a scale of 10 stars, or can make the extra effort involved in writing a comment between 10 and 1000 words.

Our dataset consists of all 1,275 USA movie titles released after year 2000 which have no less than 5,000 ratings, as of August 1, 2008.[7] For each movie we know its average rating (taken among all ratings with or without a comment), as well as the value and date-stamp of its each commented rating, but we do not have any specific information about each uncommented rating. There are 407,557 commented ratings in total. We divided the movies into 3 tiers based on their ratings: the top tier contains 545 "good movies" whose ratings are at least 7, the bottom tier contains 34 "bad movies" whose ratings are at most 4, and the middle tier contains the rest.

We observed a clear decreasing trend in the ratings associated with comments in the top tier, as shown in Fig. 6. We did not observe any clear trend in the bottom tier. Like in the Amazon study, we tested the alternative hypothesis "$\bar{X}_{l-n+1,l} > \bar{X}_{1,n}$" in each of the three tiers, for $n = 1, \ldots, 50$. The 100 tests in the top tier and the middle tier all support the alternative with p-value less than 0.001. None of the 50 tests in the bottom tier yield statistically significant results (i.e. the alternative can neither be supported nor rejected), possibly due to its small size. While it is not reliable to say that bad movies tend to receive higher ratings, it is safer to conclude that good movies accumulate lower ratings as time goes on.

We also examined the difference between the overall average rating (with or without a comment) and the average rating associated with a comment for each movie, and the result is shown in Fig. 7. Assuming as we do that the

[7] The list of titles were obtained using IMDB's Power Search engine. A constraint on the number of ratings is necessary because the proportion of commented rating among all ratings is very low (1.17% in our dataset).

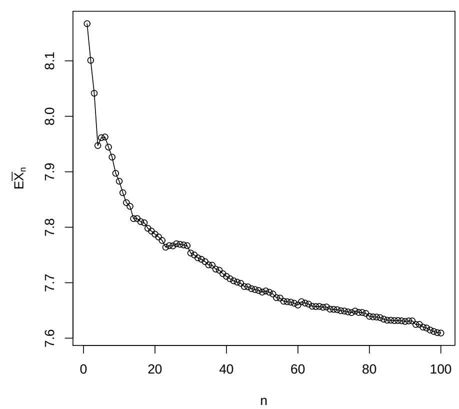


Figure 6: Average rating associated with a comment of movies with average rating no less than 7, as a function of the number of existing ratings. It can be seen that good movies tend to receive lower ratings as time goes on.

benefit of opinion expression diminishes as the number of opinions grows, it follows that it is those who are willing to incur a high cost that will continue to express opinions at very late times. Since the data in the figure shows that the latter deviate opposite from the population average, this offers a tentative (but not conclusive)[8] explanation for the moderating effect in the average opinion.

These results show that in the process of articulating and expressing their views online, people tend to follow a different pattern from that observed in information cascades or group polarization. This might be due to the fact that both at `Amazon` and `IMDB`, the setting and measurements differ from the standard ones in group polarization experiments and information cascades. What is observed instead is an anti polarization effect, whereby previous

[8] An alternative explanation is that people are reluctant to be the first one to dissent, so after the first bad review appears, a series of bad reviews follow up. To test the effect of pioneering bad reviews, we counted the number of 1-star ratings among the first n ratings of each `Amazon` title (denoted by Y_n), and then performed the linear regression $X_{n+1} \sim a + b\bar{X}_n + cY_n$ for $n = 2, 3, 4, 5$. All the fittings yield a statistically significant coefficient c between -0.002 and -0.001, suggesting a weak herding effect.

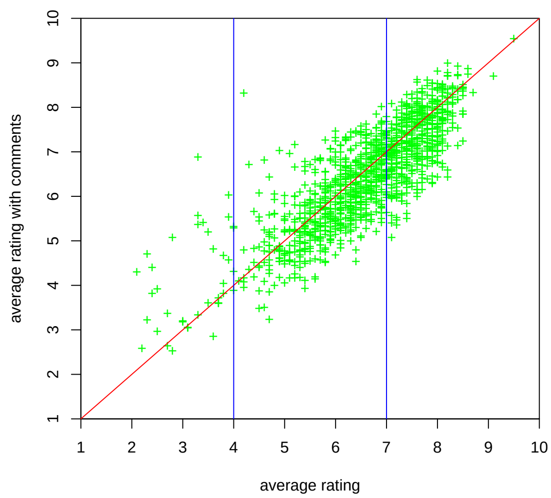


Figure 7: Selection bias induced by commented ratings. Each cross in this figure corresponds to one movie title. The horizontal coordinate represents the movie's overall average rating ($\bar{r}$) taken over both commented and uncommented ratings. The vertical coordinate represents the movie's average rating taken over only commented ratings ($\bar{r}_c$). The two vertical lines separate the data points into three tiers. Clearly, points in the bottom tier tends to fall above the red line, and points in the top tier tends to fall below. That is to say, those users who spend the additional cost to write a comment tend to express opinions opposite to those of the majority. A t-test of the alternative hypothesis that $\bar{r}_c < \bar{r}$ for good movies and a similar t-test of $\bar{r}_c > \bar{r}$ for bad movies both yield a p-value less than 0.001.

comments and ratings elicit contrarian views that soften the previous opinions.[9]. We point out however, that it is opinions like the ones we studied that are increasingly used by millions of people to navigate the sea of information they encounter every day. Thus these results might be of practical value.

In closing, these results throw a cautionary note on the interpretation of online public opinion. This is because a simple change in the order or frequency of given sets of views can change the ongoing expression in the community, and thus the perceived collective wisdom that new users will find when accessing that information.

We thank the anonymous referees for a number of constructive suggestions.

References

[1] Chrysanthos Dellarocas. Immunizing online reputation reporting systems against unfair ratings and discriminatory behavior. In *ACM EC'00*, Minneapolis, Minnesota, USA, October 2000.

[2] Dan Cosley, Shyong K. Lam, Istvan Albert, Joseph A. Konstan, and John Riedl. Is seeing believing? How recommender interfaces affect users' opinions. In *CHI 2003*, Ft. Lauderdale, Florida, USA, April 2003.

[3] Guodong Gao, Bin Gu, and Mingfeng Lin. The dynamics of online consumer reviews. In *WISE 2006*, Evanston, Illinois, USA, December 2006.

[4] Nan Hu, Paul A. Pavlou, and Jennifer Zhang. Can online reviews reveal a product's true quality? Empirical findings and analytical modeling of online word-of-mouth communication. In *ACM EC'06*, Ann Arbor, Michigan, USA, June 2006.

[9] We point out that in a website like `Jyte.com`, where it takes only one click to agree or disagree with an arbitrary claim, we did see a strong group polarization. It is possible that the latter is due to the fact that such a vote is costless compared to the opinions on `Amazon` and `IMDB`.

[5] Chrysanthos Dellarocas. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, 49(10):1407–1424, October 2003.

[6] Abhijit V. Banerjee. A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3):797–817, 1992.

[7] Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5):992–1026, 1992.

[8] Nicholas Rescher. *Predicting the Future: An Introduction to the Theory of Forecasting*. State University of New York Press, Albany, 1998.

[9] Matthew J. Salganik, Peter Sheridan Dodds, and Duncan J. Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762):854–856, February 2006.

[10] Judith A. Chevalier and Dina Mayzlin. The effect of word of mouth online: Online book reviews. *Journal of Marketing Research*, 43(3):345–354, 2006.

[11] Xinxin Li and Lorin M. Hitt. Self selection and information role of online product reviews. In *WISE 2004*, Brisbane, Australia, November 2004.

[12] Solomon E. Asch. Opinions and social pressure. *Scientific American*, 193(5):31–35, November 1955.

[13] David G. Myers. Polarizing effects of social interaction. In H. Brandstätter, J. H. David and G. Stocker-Kreichgauer (Eds.), *Group Decision Making*, 125–161, Academic Press, London, 1982.

[14] Cass R. Sunstein. The law of group polarization. *Journal of Political Philosophy*, 10(2):175–195, 2002.

[15] Dan Ariely and Jonathan Levav. Sequential choice in group settings: Taking the road less traveled and less enjoyed. *Journal of Consumer Research*, 27:279–290, 2000.

[16] Russell Spears, Martin Lea and Stephen Lee. De-individuation and group polariztion in computer-mediated communication. *British Journal of Social Psychology*, 29:121–134, 1990.

[17] Choon-Ling Sia, Bernard C. Y. Tan and Kwok-Kee Wei. Group polarization and computer-mediated communication: Effects of communication cues, social presence and anonymity. *Information Systems Research*, 13(1):70–90, 2002.

[18] Johannes A. Zuber, Helmut W. Crott, and Joachim Werner. Choice shift and group polarization: An analysis of the status of arguments and social decision schemes. *Journal of Personality and Social Psychology*, 62(1):50–61, 1992.

[19] Maha El-Shinnawy and Ajay S. Vinze. Polarization and persuasive argumentation: A study of decision making in group settings. *MIS Quarterly*, 165–198, June 1998.

[20] Anthony Downs. *An Economic Theory of Democracy*. Harper & Row, New York, 1957.

[21] William H. Riker and Peter C. Ordeshook. A theory of the calculus of voting. *American Political Science Review*, 62:25–42, 1968.

[22] Erik Ketzan. `Amazon.com` and the new democracy of opinion. `http://www.themodernword.com/pynchon/papers_ketzan1.html`, November 2002.

[23] Chrysanthos Dellarocas and Ritu Narayan. What motivates consumers to review a product online? A study of the product-specific antecedents of online movie reviews. In *WISE 2006*, Evanston, Illinois, USA, December 2006.

[24] Roger Brown. Further comment on the risky shift. *American Psychologist*, 29:468–470, 1974.

[25] David G. Myers. The polarizing effects of social comparison. *Journal of Experimental Social Psychology*, 14:554–563, 1978.