# CAN YOU TRUST ONLINE RATINGS? EVIDENCE OF SYSTEMATIC DIFFERENCES IN USER POPULATIONS

Julie Wulff
*Copenhagen University, Copenhagen, Denmark*, juliewulff@gmail.com

Daniel Hardt
*Copenhagen Business School, Frederiksberg, Denmark*, dh.itm@cbs.dk

Julie Wulff and Daniel Hardt, 2014, "CAN YOU TRUST ONLINE RATINGS? EVIDENCE OF SYSTEMATIC DIFFERENCES IN USER POPULATIONS", Proceedings of the European Conference on Information Systems (ECIS) 2014, Tel Aviv, Israel, June 9-11, 2014, ISBN 978-0-9915567-0-0
http://aisel.aisnet.org/ecis2014/proceedings/track21/9

# CAN YOU TRUST ONLINE RATINGS? EVIDENCE OF SYSTEMATIC DIFFERENCES IN USER POPULATIONS

*Complete Research*

Wulff, Julie, University of Copenhagen/Copenhagen Business School, Copenhagen, DK, jw.itm@cbs.dk

Hardt, Daniel, Copenhagen Business School, Copenhagen, DK, dh.itm@cbs.dk

## Abstract

*Do user populations differ systematically in the way they express and rate sentiment? We use large collections of Danish and U.S. reviews to investigate this question, and we find evidence of important systematic differences: first, positive ratings are far more common in the U.S. data than in the Danish data. Second, Danish reviewers tend to under-rate their own positive reviews compared to U.S. reviewers. This has potentially far-reaching implications for the interpretation of user ratings, the use of which has exploded in recent years.*

*Keywords: Online Reviews; Sentiment Analysis; Text Mining; Social Media*

## 1 Introduction

The use of online reviews has exploded in recent years, and they now play an important role in consumer choices in a broad variety of domains, including travel, entertainment and shopping (Luca 2011). As online reviews grow in importance, it is natural to critically examine their validity. In this paper, we pose the question: do user populations differ systematically in the way they determine ratings? This is an important issue, since it could potentially introduce significant biases or distortions in reviews. It is also a natural topic to investigate, since there is abundant anecdotal evidence that such differences exist.

Consider the case of Scandinavian and American reviewers. There is a persistent stereotype, according to which Scandinavians tend to be much less positive in their evaluations than Americans. This is illustrated by the following two anecdotes. In the first anecdote, a U.S. researcher gives a talk in a Scandinavian country. After the talk, the researcher is approached by an audience member, who says, “the talk was ok”. The U.S. researcher is puzzled by this, until another member of the audience explains to him that this was actually intended to express high praise. The audience member explains that it is normal in Scandinavia to use fairly neutral terms like “ok” to express very positive evaluations. The second anecdote: a student at the beginning of his graduate studies at a U.S. university has several meetings with a prominent faculty member, and is repeatedly told that his research ideas are “wonderful”. The student is very gratified by this, until he overhears other students talking about how this faculty member seems to always respond to ideas by calling them “wonderful”. In this case, it seems that the U.S. faculty member had a fairly neutral opinion, but expressed it in a very positive way.

Of course such anecdotes do not in themselves establish the existence of systematic differences. But consider the effect of such hypothesized differences: they could be the source of significant distortion in reviews, causing users to be misled in their decisions about shopping, travel and entertainment. For example, if Scandinavians tend to give lower ratings, a given hotel might receive lower ratings because it happens to be frequented by Scandinavians; another hotel, frequented by Americans, might receive higher ratings. Such systematic differences across user communities could to a certain extent invalidate the kind of comparisons that are central to the use of online reviews.

According to the stereotype and anecdotes, it would seem that Scandinavians downgrade their positive expressions of sentiment, or, equivalently, U.S. reviewers upgrade their positive expressions. But is this stereotype actually true? In this paper, we investigate this question by analyzing large collections of online reviews by Danish and U.S. users. These reviews are short pieces of text, combined with a numerical rating which expresses the user's overall evaluation. In our view, such data should provide a meaningful test of the stereotype -- if Scandinavians and Americans do indeed differ as we have described, this should be reflected in distributional differences in these datasets.

In examining this hypothesis, we restrict attention to very positive evaluations: compared to U.S. reviewers, we expect a Danish tendency to “downgrade” from very positive to somewhat less positive. We will examine this hypothesis from two different perspectives, in looking at the positive Danish reviews vs. the positive U.S. reviews:

- **Ratings Hypothesis**: The positive Danish reviews contain relatively fewer high ratings than the positive U.S. reviews.
- **Ratings vs. Text Hypothesis**: there fewer high ratings for texts of a given positivity in the positive Danish reviews, compared to the positive U.S. reviews.

In what follows, we begin with a description of the data sets. Next we examine the distribution of ratings, to test the Ratings Hypothesis. Then we look at the text positivity: we develop a metric for positivity of terms, and examine their relative distributions. This is followed by an examination of the relation between ratings and texts in the two data sets, to test the Ratings vs. Text Hypothesis. We show that both hypotheses are strongly confirmed by the data. Finally, we observe that these results could have far-reaching implications for the interpretation of online user ratings, the use of which has exploded in recent years.

## 2 Data

In this work four datasets containing user rated reviews are presented; two of them are downloaded from Danish websites, and two from American websites. We selected reviews in two different domains: film reviews and restaurant reviews.

The Danish film data was downloaded in November 2011 from the Danish movie website *scope.dk.* This dataset contains all the rated user reviews available at the time of download: 829 films are reviewed, and the reviews total 1,624,049 words. The U.S. film data was downloaded in January 2012 from The Internet Movie Database (imdb.com) and contains rated user reviews from 678 films and has a total size of 34,599,486 words[1]. A search function on www.imdb.com was used to create a list of films and matching IMDb ID tags for films produced in the years 1920-2011. Reviews were selected only for films that were also reviewed on the Danish site. The IMDb ID tags was used to find the page containing data for each of the films and all reviews which had a correlated rating were downloaded for those 678 films. The U.S. IMDb reviews are rated on a scale of 1 to 10, while the Danish Scope reviews are rated on a scale of 1 to 6.

[1] Data from IMDb has be gathered in the movie-review-data at https://www.cs.cornell.edu/People/pabo/movie%2Dreview%2Ddata/. This data has been used for research in automatic sentiment analysis. Representative examples include Pang and Lee (2004) and Taboada, Maite, et al. (2011).

The restaurant reviews are collected from the review site *Yelp*, which has both a Danish and an American website. The Danish restaurant reviews were downloaded January 2013 from *www.yelp.dk*. All available reviews for restaurants in Copenhagen were selected, which resulted in a collection of 3,851 reviews containing 581,713 words. The U.S. restaurant reviews were downloaded in August 2012 from *www.yelp.com* and were restricted to be from restaurants in Philadelphia. This resulted in 109,129 reviews with a total of 15,161,700 words. Philadelphia was chosen as a suitable city for comparison with Copenhagen, based on size and infrastructure. Reviews from both Yelp websites are rated on a scale of 1 to 5.

## 3 Ratings

Figure 1 shows the number of reviews in each category for the U.S. film data. Here, the top category of 10 has by far the most reviews, with a ratio of 3.7 to the number of reviews for the lowest category. For the most part the number of reviews decreases as the category lowers, with a modest increase in the number of reviews for the lowest category, 1. This distribution makes intuitive sense -- it's not surprising that people would be most motivated to write reviews of films they are most enthusiastic about, and, to a lesser extent, also be motivated in cases where they have strong negative feelings. This has been noted in the literature: Wu and Huberman (2010) point out that the so-called "brag and moan" view of ratings is fairly typical (as also mentioned by Hu et al. (2006) and Dellarocas and Narayan (2006)). The tendency of the top category to be the most frequent is also mentioned on the yelp.com site, where the top category of 5 is claimed to be the most frequent: "The numbers don't lie: people love to talk about the things they love!" (yelp.com (2012) )

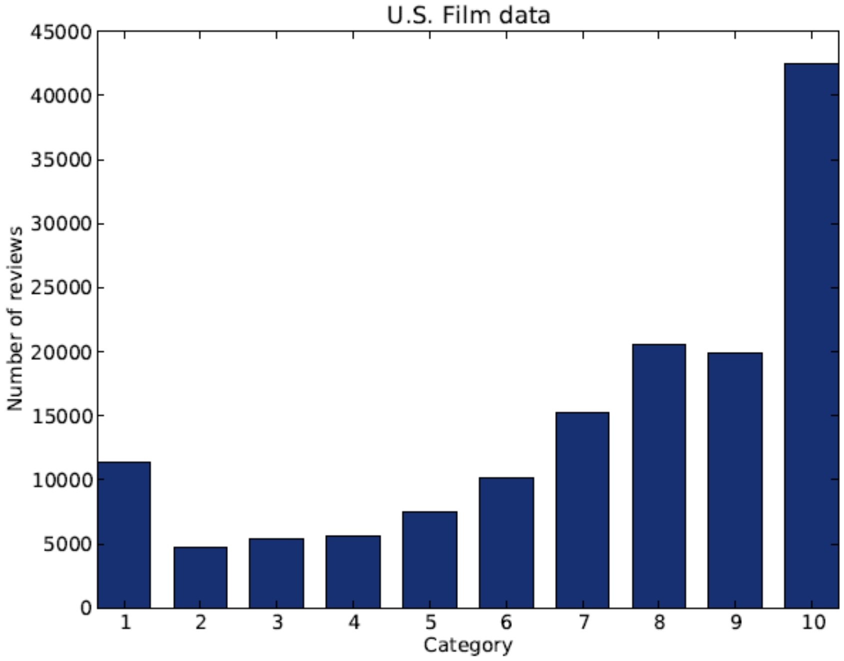

*Figure 1. U.S. film reviews (IMDb) per category.*

There is a very different distribution in the Danish Scope data, as shown in Figure 2. Here, category 4 (out of 6) is the most frequent with around one third more reviews than the highest category. This provides dramatic support for the Ratings Hypothesis: highly positive evaluations are over-represented in the U.S. data compared to the Danish data.

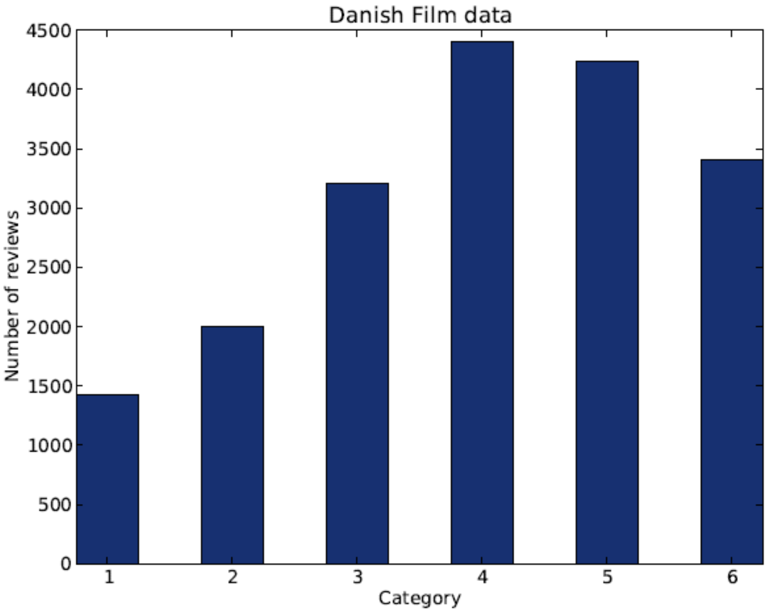

*Figure 2. Danish film reviews (Scope) per category.*

We turn now to restaurant reviews. While the distribution is somewhat different, it also strongly supports the Ratings Hypothesis. It can be observed in Figure 3 that U.S. restaurants reviews have an overrepresentation of highly rated reviews. Unlike the film reviews, here the U.S. data actually shows a small decrease in the number of reviews in the top category. A second difference is that there is no increase in the number of reviews for the lowest category, as was observed in the U.S. film data. This might perhaps be related to differences between the film domain vs. the restaurant domain – perhaps films elicit more enthusiasm in both the positive and negative direction, so that, in the U.S. data, the so-called "brag and moan" pattern fits better to film reviews than restaurant reviews. This makes intuitive sense – a film review can be thought of as an expression of an aesthetic judgement and is in that sense perhaps somewhat unconstrained by practical considerations. A restaurant review, by contrast, involves other, more practical factors in addition to purely aesthetic considerations. For example, if a reviewer had a very positive experience in terms of the food itself, this would push the reviewer towards a very positive rating, but the reviewer might also feel somewhat compelled to consider practical considerations such as price, the restaurant location, and the service. We imagine that aesthetic judgements might be more likely to result in extremely positive or negative ratings than judgements concerning practical considerations. We won't pursue these speculations further here, but we note that the evident differences between film and restaurant reviews makes it even more striking that they both strongly support our hypothesis about the differences between Danish and U.S. ratings.

This can be observed by comparing Figure 3 with Figure 4. Just as with the film data, we see that the distribution of reviews per category is shifted to the left for the Danish data, as compared to the U.S. data. More specifically, the relative number of reviews in the top category is much lower in the Danish data than in the U.S. data. This can be quantified by examining the ratio of the number of reviews in the top category to the number in the next highest category. In the Danish film data there are 3403 reviews in the top category vs. 4242 reviews in the next highest category, for a ratio of .80. The U.S. film data has 42441 reviews in the top category vs. 19942 in the next category, for a ratio of 2.13 – the ratio in the U.S. data is nearly three times the Danish ratio. Similarly in the restaurant data the corresponding Danish ratio is 334 vs. 1387, or .24, while the U.S. restaurant ratio is 37747 vs. 41516, or .91 – over three times the Danish ratio.

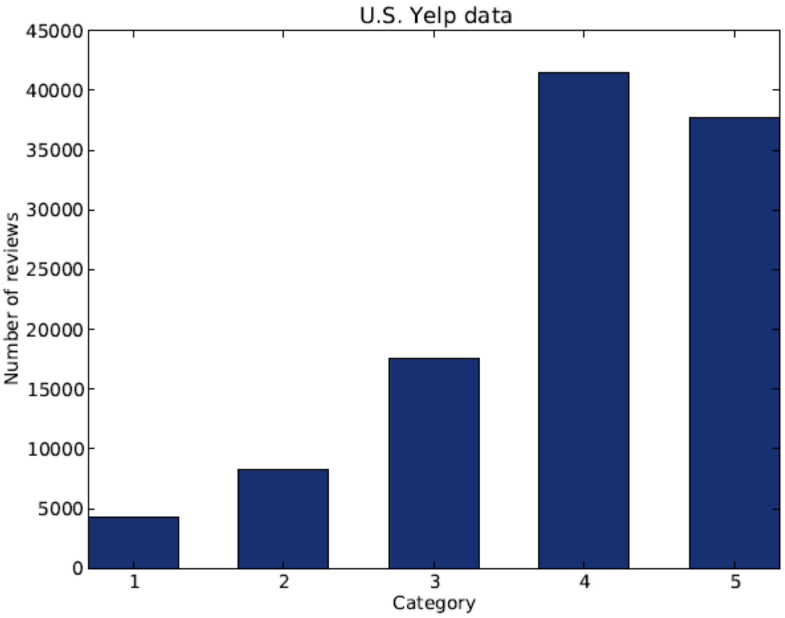

*Figure 3. U.S. restaurant reviews (Yelp) per category.*

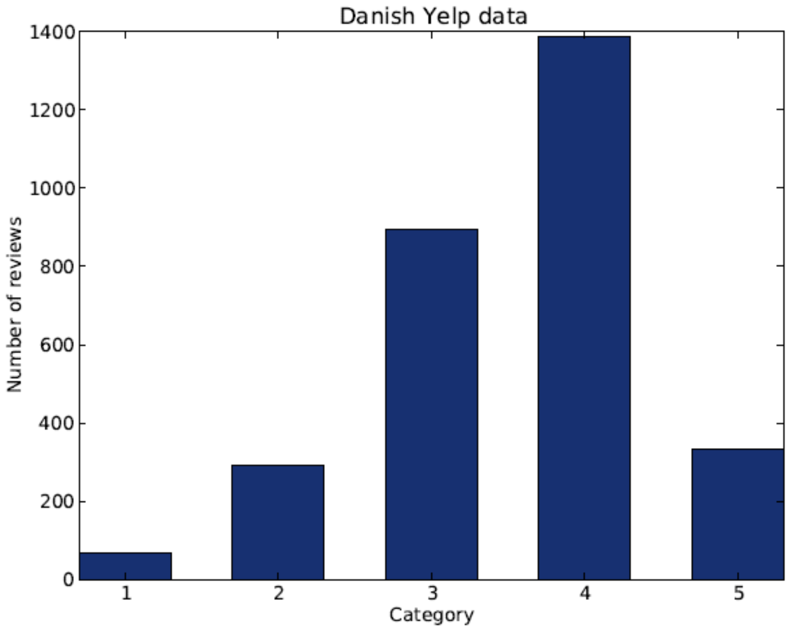

*Figure 4. Danish restaurant reviews (Yelp) per category.*

# 4 Text

We have seen that the Ratings Hypothesis is strongly confirmed: there are fewer highly positive ratings in the Danish data than in the U.S. data. We now wish to examine the Ratings vs. Text

Hypothesis: that there would be fewer highly positive Danish ratings for texts of a given positivity. To examine this, we first need a way to measure the positivity of a text. We employ techniques from the field of sentiment analysis (Pang and Lee 2008) for this purpose.

A standard simplifying assumption, discussed at length by Pang and Lee, is that the sentiment of a text can be assessed by examining the words and terms occurring in it. We follow that assumption here, and thus we begin by computing the occurrences of terms in each text, and produce a positive/negative sentiment lexicon containing those terms. In this paper we consider a term to be a short sequence of 1-3 words. To avoid bias in the data, we only consider terms that occur in reviews from more than one film or restaurant. The data contains reviews rated by three different rating scales, 1-5 stars, 1-6 stars and 1-10 stars. We assume the rating scales to be continuous and normalize rating scales for all four datasets by rescaling ratings to run from -0.5 to 0.5. Intuitively, highly positive terms are those most frequent in the top category and most infrequent in the other categories. We measure the frequency of a term per category and use it to find a value reflecting the positivity of a term. We calculate the expected category for a given term, which is a weighted average of normalized frequencies and thereby provides the best guess of a category for a given term, and use this as a positivity measurement.

$$Frequency = \frac{Count_n}{Total_n}$$

where $Count_n$ is the number of times a given term with length $n$, occur in a category and $Total_n$ is the total number of terms with length $n$ in that category.

$$Normalized\ Frequency = \frac{Frequency}{\sum_1^i \frac{Count_n}{Total_n}}$$

where $i$ is the number of categories.

$$Expected\ Category = \sum_1^i Normalized\ Frequency \times Category$$

where $Category$ is the rescaled value for category $i$.

These measurements deliver a sentiment scale for all terms with continuous values in the range -0.5 to 0.5, with -0.5 representing the most negative term and 0.5 the most positive.[2] Manual inspection suggests that this technique indeed provides a faithful ordering of terms based on their positivity. In Tables 1 through 4, we give an extraction from the sentiment lexicon, showing the top 24 most negative and positive terms for both U.S. and Danish film data. The most negative or positive terms appear at the top of the list.

[2] See http://sentiment.christopherpotts.net for detailed descriptions of these and similar relevant techniques.

**Table 1: 24 Most Negative Terms IMDb**

| Negativity | Term |
|---|---|
| -0.4695 | awful movie this |
| -0.4671 | the worst piece |
| -0.4656 | 10 worst |
| -0.4651 | absolutely no redeeming |
| -0.4587 | 1 of 10 |
| -0.4583 | horrible piece of |
| -0.4581 | horrible piece |
| -0.4571 | describe how bad |
| -0.4567 | worst piece of |
| -0.4566 | awful movie ! |
| -0.4566 | worst piece |
| -0.4543 | crap !!! |
| -0.4531 | <s> awful ! |
| -0.4531 | avoid ! </s> |
| -0.4502 | this garbage , |
| -0.4502 | piece of dreck |
| -0.4502 | money back after |
| -0.4501 | ever walked out |
| -0.4496 | this worthless |
| -0.4495 | this laughable |
| -0.4491 | <s> !!! </s> |
| -0.4487 | money back ! |
| -0.4408 | 0 stars |
| -0.4466 | ... the worst |

**Table 2: 24 Most Positive Terms IMD**

| Negativity | Term |
|---|---|
| 0.4943 | ! 10 |
| 0.4932 | ! 10 / |
| 0.4839 | . 10 out |
| 0.4754 | masterpiece !!! |
| 0.4742 | $<s>$ perfection |
| 0.4728 | a ++ |
| 0.4720 | is my absolute |
| 0.4692 | superb script |
| 0.4679 | than his father |
| 0.4678 | 2nd favourite |
| 0.4668 | 11 out of |
| 0.4668 | 11 out |
| 0.4647 | has changed my |
| 0.4640 | ... 10 / changed my |
| 0.4634 | then i strongly |
| 0.4633 | 5 best movies |
| 0.4633 | love , death |
| 0.4633 | this incredible movie |
| 0.4628 | than a 10 |
| 0.4628 | yes yes yes |
| 0.4626 | loved everything about |
| 0.4621 | moves me |
| 0.4607 | brings tears to |
| 0.4606 | , great writing |

**Table 3: 24 Most Negative Terms Scope**

| Negativity | Term |
|---|---|
| -0.4768 | dårligste film jeg (worst movie I) |
| -0.4766 | , plat (, lame) |
| -0.4755 | lorte film (shitty movie) |
| -0.4745 | dårlig en (bad one) |
| -0.4677 | ret elendig (pretty poor) |
| -0.4652 | ringeste (the worst) |
| -0.4604 | ikke er værd (not worth) |
| -0.4604 | ringe , at (poor , to) |
| -0.4600 | dine penge (your money) |
| -0.4596 | at spilde (to waste) |
| -0.4591 | gang lort (some crap) |
| -0.4569 | makværk <s> . (mess <s> .) |
| -0.4565 | makværk .(mess .) |
| -0.4564 | noget bras (some junk) |
| -0.4534 | ligegyldig film |
| -0.4528 | dårligt |
| -0.4508 | dragen (dragon) |
| -0.4507 | makværk (mess ) |
| -0.4493 | de dårligste film (the worst movies) |
| -0.4493 | talentløs (talentless) |
| -0.4490 | dum film (stupid movie) |
| -0.4490 | Intet fungerer (nothing works) |
| -0.4467 | en <s> spild (a waste) |
| -0.4450 | min tid (my time) |

**Table 4: 24 Most Positive Terms Scope**

| Negativity | Term |
|---|---|
| 0.4781 | elsk (love) |
| 0.4841 | verdens bedste film (the world's best movie) |
| 0.4781 | ret den bedste ! (the best !) |
| 0.4738 | go og (go and) |
| 0.4732 | ret mesterværk (masterpiece) |
| 0.4657 | Mesterværk anonym (masterpiece anonymous) |
| 0.4630 | ret kanon (pretty great) |
| 0.4613 | bedste film (best movie) |
| 0.4609 | bedste tegnefilm (best cartoon) |
| 0.4608 | får 6 (get 6) |
| 0.4608 | elsker bare (just love) |
| 0.4603 | ses ! <s> (watch) |
| 0.4600 | ses ! (watch) |
| 0.4589 | kan se igen (can see again) |
| 0.4589 | go ! <s> (good) |
| 0.4567 | støreste film (biggest movie) |
| 0.4559 | så smuk (so beautiful) |
| 0.4548 | skal se ! (must see) |
| 0.4548 | får 6 stjerner (get 6 stars) |
| 0.4521 | Bedste film jeg (best movie I) |
| 0.4515 | utrolig rørende (unbelievably touching) |
| 0.4507 | jeg elsker bare (I simply love) |
| 0.4507 | den er super (it is super) |
| 0.4503 | eneste film (the only movie) |

# 5 Ratings vs. Text

Now that we have a measure of the positivity of the text, we are in a position to examine the Ratings vs. Text Hypothesis, namely, that a text of a given positivity will get a lower rating in the Danish data than in the U.S. data. As described in Section 4, we calculate the positivity of a text as the average of the expected category values for terms in the text. From these calculations a distribution of positivity over categories is achieved, for each dataset. We compare the positivity both for Danish and U.S. data in the two domains. Figure 5 shows the result of our positivity calculations for the Danish and U.S. film dataset and Figure 6 shows the results for the restaurant datasets. As one would expect, positivity is strongly correlated with rating categories in all datasets. However there is a striking difference in the top rating categories. This difference concerns the slope of the line that maps category to text positivity. The slope is noticeably steeper in the Danish data as it moves towards the top category. This can be clearly seen in both the film data and restaurant data.

To see what this means, consider that, in general, text positivity increases as the rating category increases. The slope of the line provides a measure of how much text positivity must increase to support a change in rating category. What we have observed is that, near the top rating categories, U.S. data does not require as much increase in text positivity as the Danish data – in other words, a U.S. reviewer is more willing to give a top rating for a text of a given positivity than a Danish reviewer would be. Interestingly, in the middle area the slopes are identical in the Danish and U.S. data, while a difference is also observed in the lower categories. So this difference concerns the ratings for the most positive and most negative categories. Finally, it is observed both the film and the restaurant domains. This is particularly striking in view of the fact that there's reason to believe that film and restaurant ratings differ in many ways.

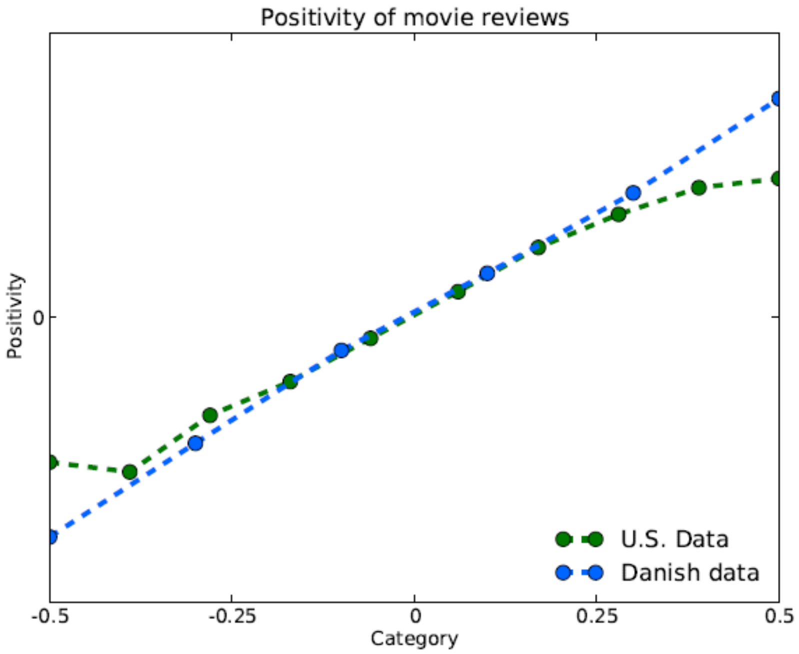

*Figure 5. Positivity of the text in the two film datasets across categories.*

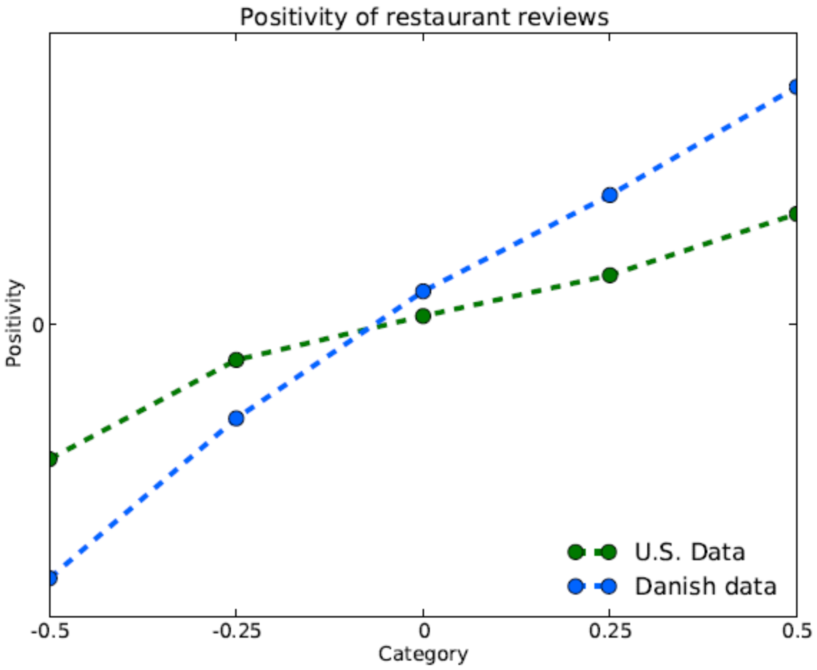

*Figure 6. Positivity of the text in the two restaurant datasets across categories.*

This analysis strongly supports the Ratings vs. Text Hypothesis: for positive reviews of a given positivity, the ratings tend to be higher in the U.S. data. In other words, Danish reviewers, when compared to U.S. reviewers, have a tendency to "downgrade" their positive reviews.

## 6 Conclusion

There is a widely held belief that Americans and Scandinavians differ in the way they express and rate positive sentiment. To our knowledge this paper represents the first attempt to test such a belief in a systematic way. We have expressed this hypothesis in terms of the Ratings Hypothesis and the Ratings vs. Text Hypothesis. Using large collections of Danish and U.S. online reviews, we have found strong confirmation for both hypotheses, in two very different domains, films and restaurants.

In recent years, the use of rating systems have exploded, to the point where they are relied on every day for millions of decisions about everything from where to eat to what film to see, or where and how to take a vacation. The present work, while limited in scope, suggests a potentially far-reaching conclusion; namely, it points to the possibility that there are systematic differences in rating systems, that we ignore at our peril. As we have seen, Danes differ sharply from Americans in the positivity of ratings and text: they give far fewer top ratings; and they tend to give lower ratings for texts of a given positivity. One natural conclusion is that there are cultural differences leading Danes to produce reviews and ratings in a rather different way than Americans. In our experience, those familiar with Danish and American culture find this quite plausible and readily suggest numerous potential explanations -- perhaps the most compelling of which concerns the traditional grading system in Danish schools[3], where the top grade of "13" was given in only the most exceptional of circumstances, and was always far less frequent than the top grade of "A" in U.S. schools.

One alternative explanation for these differences would be to appeal to differences in the domains being evaluated by the reviewers. For example, in the film domain, it could be that Danes are simply less enthusiastic about the films they see. This might seem somewhat paradoxical -- since Danes and Americans are both free to choose which films they see, one might expect that they are equally enthusiastic about the films they choose to see and review. However, it has often been suggested that the film industry in many European countries is subject to U.S. cultural imperialism, which would hold that, because of its economic and cultural power, the U.S. film industry is able to substantially alter the film-going options of the Danish public.

In our view, this explanation loses whatever plausibility it might have in view of the fact that we have found similar effects in a second domain, namely restaurants. In both domains, there are significantly more top ratings in the American data than in the Danish data, and it is difficult to see why Danes should in general be more negative about both the films they see and the restaurants they attend. Furthermore, it is striking that, in both domains, we find a systematic differences between Danes and Americans for texts expressing a similar level of positivity -- Danes tend to move many of these from a top category to a less positive one. This was found by examining the slope of the line as it moves to the higher categories; in both the restaurant and film domain, the effect was strikingly similar. In our view this constitutes clear evidence of a systematic difference in how ratings are produced by these two populations.

We have argued that these differences point to a potentially important problem with the use of rating systems, especially if such differences are widespread. In future work, we intend to examine reviews in other domains, to see if the differences we have found are consistent across different domains. In addition, we plan to examine other user communities to see if similar systematic differences can be found. We are also exploring ways to address the problem these differences pose: one natural hypothesis is that, when there is a systematic mismatch between text and rating, the text positivity is a better guide to the true sentiment. We would like to see if an automatic sentiment analysis might reduce systematic mismatches in these cases.

## References

C. Dellarocas and R. Narayan (2006). What motivates consumers to review a product online? A study of the product-specific antecedents of online movie reviews. In Proceedings of the International Conference on Web Information Systems Engineering.

Noah Constant, Christopher Davis, Christopher Potts and Florian Schwarz. (2009). The pragmatics of expressive content: Evidence from large corpora. Sprache und Datenverarbeitung, 1 (2), 5–21.

Wenjing Duana, Bin Gub, Andrew B. Whinston (2008). Do online reviews matter? — An empirical investigation of panel data. Decision Support Systems. Volume 45, Issue 4, November 2008, Pages 1007–1016

FAQ (2012). yelp.com. http://www.yelp.com/faq.

N. Hu, P. A. Pavlou, and J. Zhang (2006). Can online reviews reveal a product's true quality? Empirical findings and analytical modeling of online word-of-mouth communication. In Proceedings of the ACM Conference on Electronic Commerce.

[3] The Danish grading system was revised in 2006, in part to make it more in line with grading systems in other countries (Wikipedia: Academic grading in Denmark, 2012).

Michael Luca (2011). Reviews, reputation, and revenue: The case of Yelp.com. Harvard Business School.

Pang, B., and Lee, L.(2004). "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts." *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 2004.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, *2*(1-2), 1-135.

Christopher Potts (2012). Extracting social meaning and sentiment. NASSLLI 2012: http://nasslli2012.christopherpotts.net/ratings.html.

Qiang Yea, Rob Lawb, Bin Guc, Wei Chen (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings Computers in Human Behavior, Volume 27, Issue 2, March 2011, Pages 634–639

Wikipedia (2012). Academic grading in Denmark. [Online; accessed 27-August-2012].

Fang Wu and Bernardo Huberman (2010). Opinion formation under costly expression. ACM Transactions on Intelligent Systems and Technology, 1(1).

Qiang Yea, Rob Lawb, Bin Guc and Wei Chen (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. Computers in Human Behavior, 27 (2), 634–639.

Taboada, Maite, et al. (2011) "Lexicon-based methods for sentiment analysis." *Computational linguistics* 37.2: 267-307.