# Broader implications: sociology/social psychology

**What opinions are influential?**
    ⟶ proxy question: which Amazon reviews are rated helpful?
    [Danescu-Niculescu-Mizil, Kossinets, Kleinberg, and Lee '09]

Prior work has focused on features of the *text* of the reviews, and has not been in the context of sociological inquiry. [Kim et al. '06, Zhang and Varadarajan '06, Ghose and Ipeirotis '07, Jindal and B. Liu '07, J. Liu et al '07, Otterbacher '09 (who also considers some metadata features)].

Our focus: how about *non-textual* features (social aspects, biases)?

Our corpus: millions of Amazon book reviews.

## Some social factors boosting helpfulness scores

- using "real name"

## Some social factors boosting helpfulness scores

- using "real name"
- being from New Jersey (for science books)

# Some social factors boosting helpfulness scores

- using "real name"
- being from New Jersey (for science books)
- not being from Guam
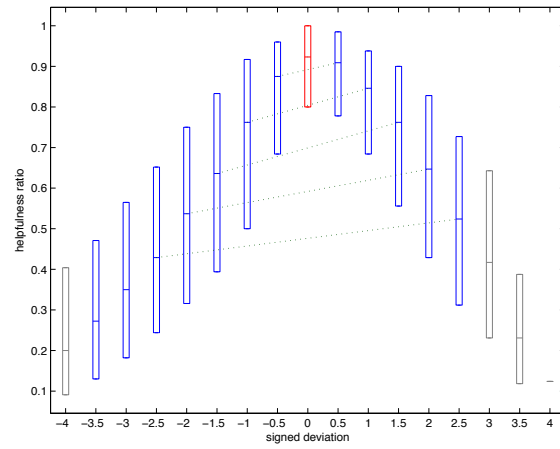
## Some social factors boosting helpfulness scores

- using "real name"
- being from New Jersey (for science books)
- not being from Guam

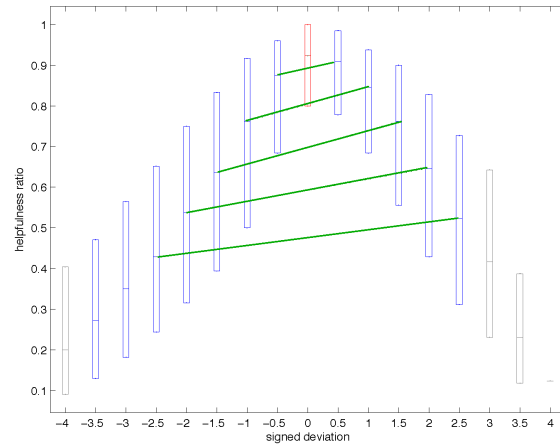Our focus: What about the review's star rating in relationship to others?

Theories from social psychology:

- conform (to the average rating) [Bond and Smith '96]
- "brilliant but cruel" [Amabile '83]
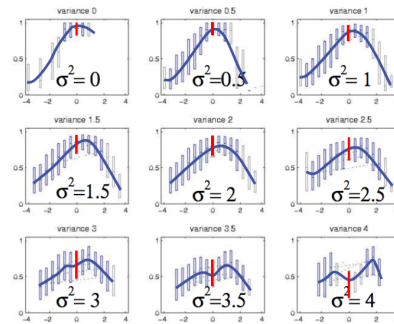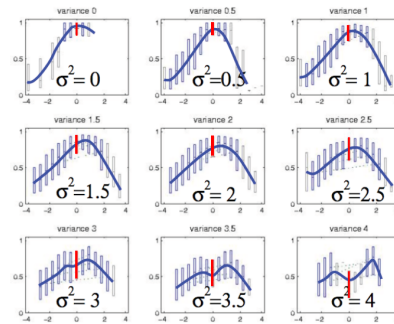
# Best to conform

# Best to conform...?

# New observation: effect of variance

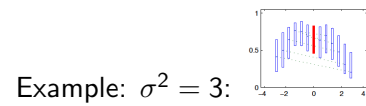As *variance* among reviews increases, be *slightly above* the mean

# New observation: effect of variance

As *variance* among reviews increases, be *slightly above* the mean



... **except in Japan**, where it's best to be *slightly below*.



Example: $\sigma^2 = 3$:

# Are the social effects just textual correlates?

**We would like to control for the actual quality of a review's text.** (Maybe NJ people really do write better reviews?)

How should we determine the "real" helpfulness, in order to control for it?

- ▶ manual annotation? Tedious, subjective.
- ▶ automatic classification? Need extremely high accuracy guarantees.

## Are the social effects just textual correlates?

**We would like to control for the actual quality of a review's text.** (Maybe NJ people really do write better reviews?)

How should we determine the "real" helpfulness, in order to control for it?

- ▶ manual annotation? Tedious, subjective.
- ▶ automatic classification? Need extremely high accuracy guarantees.

It turns out that 1% of Amazon reviews are *plagiarized*! (see also David and Pinch ['06]).

Our social-effects findings regarding position relative to the mean hold on plagiarized pairs, which *by definition* have the same textual quality.

7 of 11 people found the following review helpful:

★☆☆☆☆  **Migraine Headache at No Extra Charge**, May 28, 2004

By **A Customer**

If you enjoy a thumping, skull splitting migraine headache, then the Sing N Learn series is for you.

As a longtime language instructor, I agree with the effort that this series makes, but it is the execution that ultimately weakens Sing N Learn series. To be sure, there are much, much better ways to learn a foreign language. In fact, I would recommend this title only as a last resort and after you've thoroughly exhausted traditional ways to learn Korean.

Amazon.com: Customer Reviews: Sing 'n Learn Chinese (Book & CD)                                                                11/1/08 8:45 PM

26 of 30 people found the following review helpful:

★☆☆☆☆  **Skull-splitting headache guaranteed!!**, June 16, 2004

By **A Customer**

If you enjoy a thumping, skull splitting migraine headache, then Sing N Learn is for you.

As a longtime language instructor, I agree with the attempt and effort that this series makes, but it is the execution that ultimately weakens Sing N Learn Chinese.

To be sure, there are much, much better ways to learn Chinese. In fact, I would recommend this title only as a last resort and after you've thoroughly exhausted traditional ways to learn Chinese.

## Read the book

Even more information about applications, research directions, connections to other fields, and snazzy examples is available:

*Opinion Mining and Sentiment Analysis*
Bo Pang and Lillian Lee
www.cs.cornell.edu/home/llee/opinion-mining-sentiment-analysis-survey.html
Now Publishers (also Foundations and Trends in Information Retrieval 2(1-2), 2008

135 pp, 330+ references, full pdf posted and freely available online

## Our revels now are ended

We have seen that sentiment analysis...

...has many important applications
...encompasses many interesting research questions
...extends to many areas

## Our revels now are ended

We have seen that sentiment analysis...

...has many important applications
...encompasses many interesting research questions
...extends to many areas

You (and students in your classes) can start working on sentiment
analysis right now!
Publicly available datasets include those at:

http://www.cs.cornell.edu/home/llee/data

## Our revels now are ended

We have seen that sentiment analysis...

...has many important applications
...encompasses many interesting research questions
...extends to many areas

You (and students in your classes) can start working on sentiment analysis right now!
Publicly available datasets include those at:

http://www.cs.cornell.edu/home/llee/data

This is such stuff as dreams are made on!