Social Media Brand Tracking

Wendy W. Moe
University of Maryland

Social Media Monitoring Research

- User behavior has implications for data and metrics
  - Individual level analysis to understand posting incidence and evaluation behavior
    - Moe, Wendy W. and David A. Schweidel (2012), "Online Product Opinion: Incidence, Evaluation and Evolution," Marketing Science, 31 (3), 372-386
  - Product level analysis to examine the impact on sales
    - Moe, Wendy W. and Michael Trusov (2011), "The Value of Social Dynamics in Online Product Ratings Forums," Journal of Marketing Research, 48 (3), 444-456.
  - Antecedents and Consequences of Engagement
    - Weiger, Welf H., Wendy W Moe, Hauke A. Wetzel, and Maik Hammerschmidt (2016), "Behavioral Engagement in Social Media: Measurement, Antecedents, and Purchase Consequences," *Working Paper.*

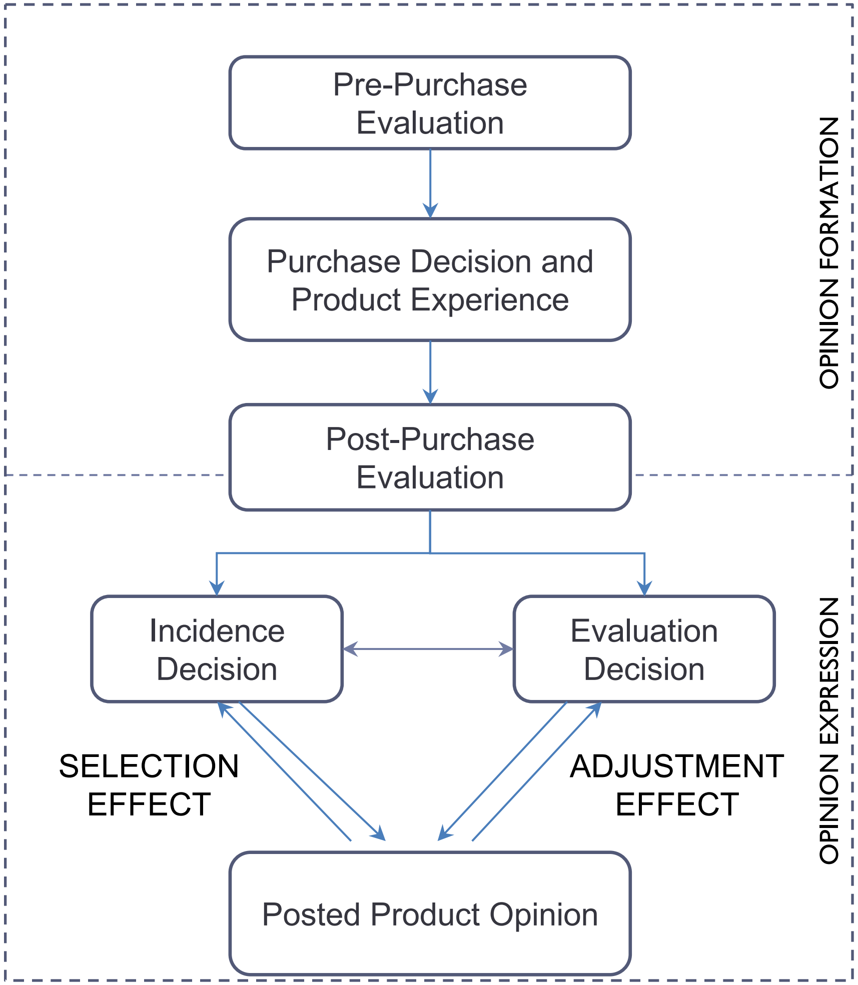
- Brand tracking over time using social media data can potentially mirror traditional offline marketing research
  - Schweidel, David and Wendy W. Moe (2014), "Listening in on Social Media: A Joint Model of Sentiment and Venue Format Choice" Journal of Marketing Research, 51 (4), 387-402.

- Can social media data help brands benchmark to other brands?
  - Zhang, Kunpeng and Wendy W. Moe (2016), "A Social Media Based Method for Measuring Brand Favorability while Accounting for User Scale Usage Heterogeneity," *Working Paper.*

Individual user behavior (Moe and Schweidel 2012*)

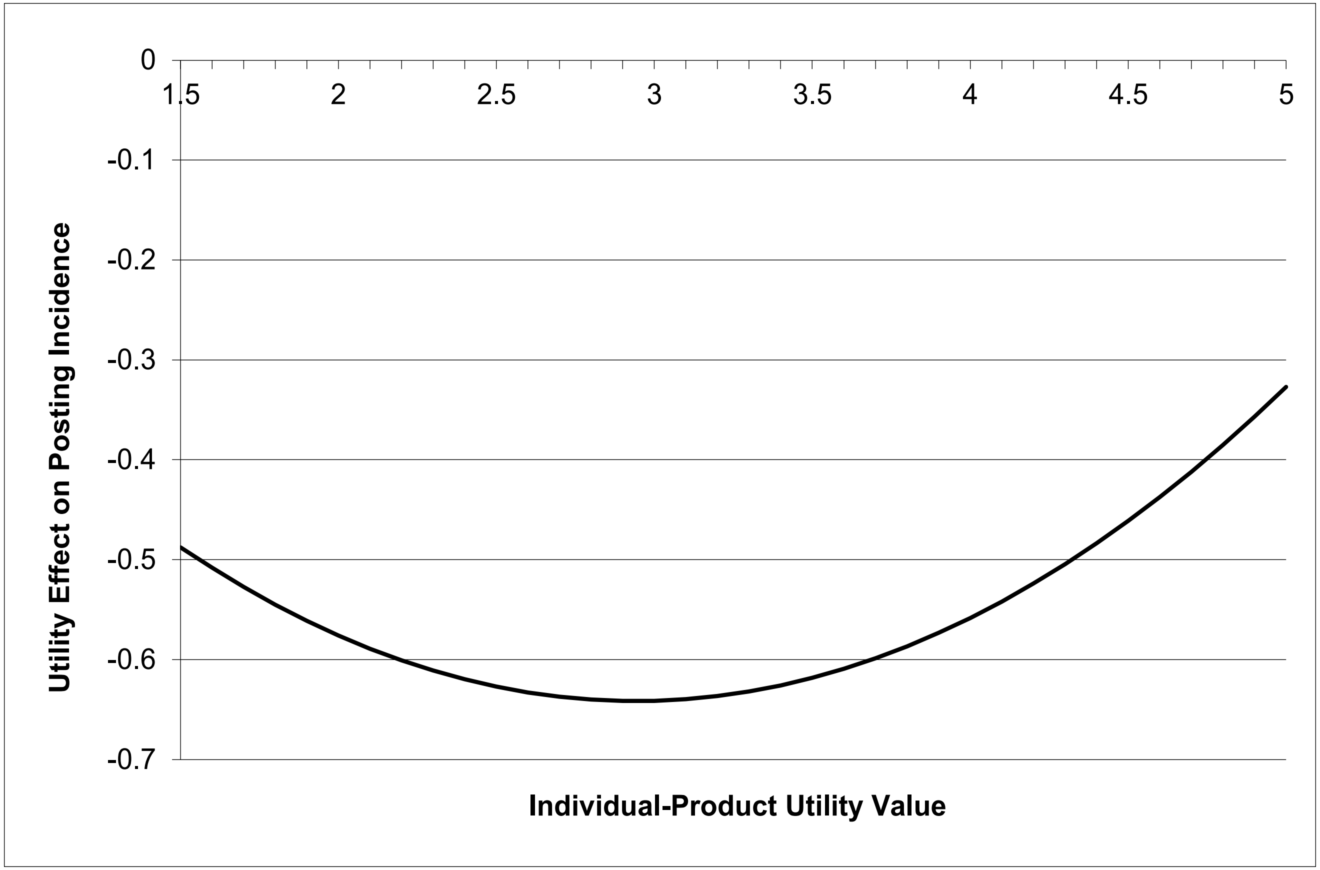
What influences posting behavior?

- Opinion formation vs. opinion expression
- Opinion formation, in theory, is a function of satisfaction
- Opinion expression is subject to a variety of biases and dynamics
  - Scale usage
  - Expert effects
  - General audience effects
  - Multiple audience effects
  - Bandwagon vs. differentiation
- Example: Opinion polls and voter turnout

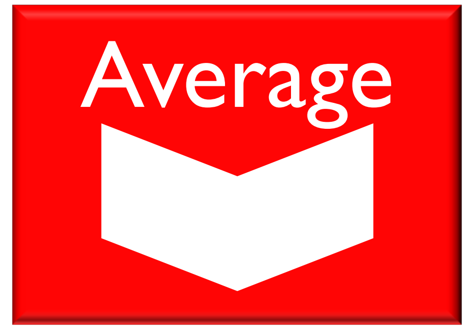


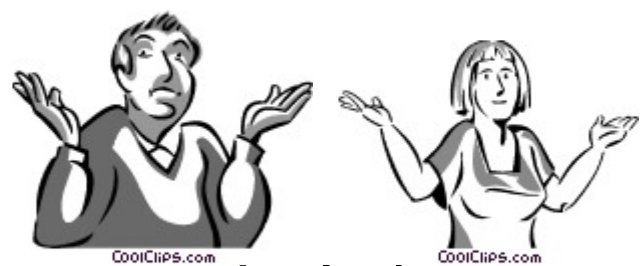
* Data from product ratings posted to an retailer's online site.

Selection Effect I: Role of Underlying Opinion



Selection Effect II:
Social Dynamics over Time





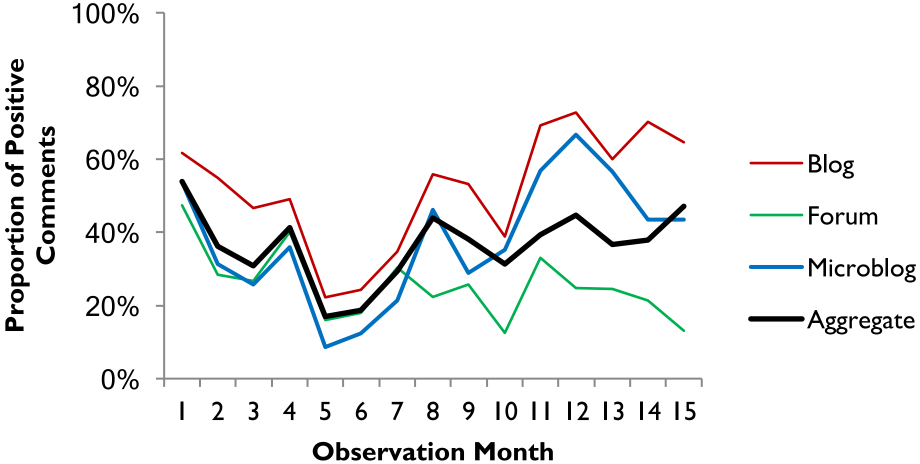
<u>Activists</u>

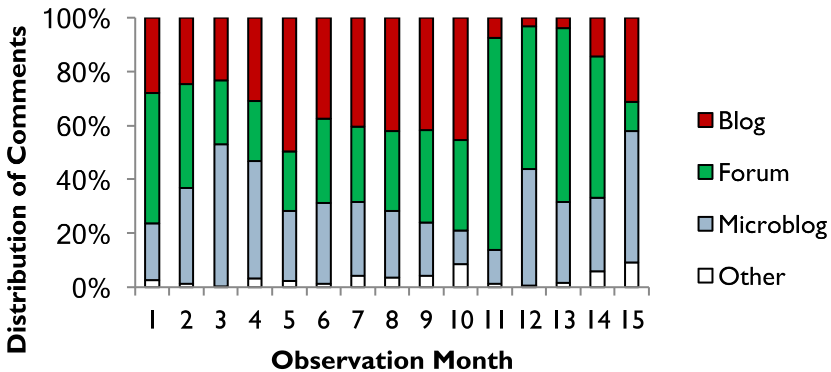
<u>Low Involvements</u>

Post frequently
Attracted by lack of consensus
More negative
Variance and volume make them more negative

Post infrequently
Deterred by lack of consensus
More positive
Variance and volume make them more positive

Brand Monitoring over Time (Schweidel and Moe 2014*)





Correlation with offline brand tracking survey

Venue	Correlation
Blogs	.197
Forums	-.231
Microblogs	-.394
Average	.008

* Data from social media monitoring of an enterprise software brand.

What factors influence social media sentiment?

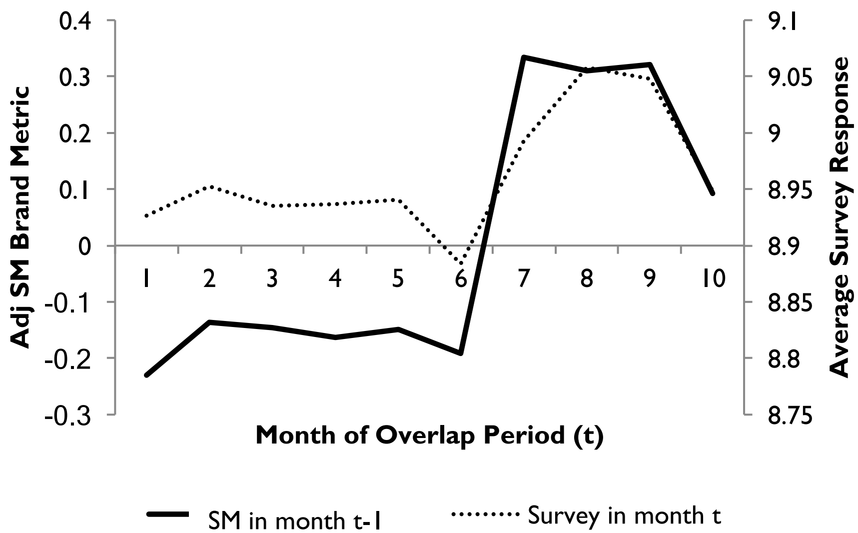
General Brand Impression (GBI)
(Dillon et al 2001)

Venue

Venue-specific dynamics

Message topic

Potential to be an early indicator

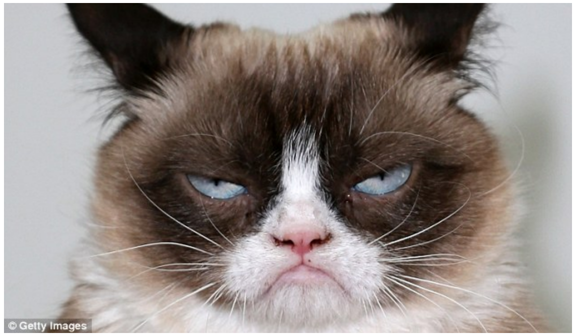


- Correlation with survey (t)
  - Adj SM Brand Metric = .376
  - Avg sentiment =.008
  - Blogs = .197
  - Forums = -.231
  - Microblogs = .394
- Correlation with survey (t+1)
  - Adj SM Brand Metric = .881
  - Avg sentiment = .169
  - Blogs = .529
  - Forums = .213
  - Microblogs = .722

from Schweidel, David and Wendy W. Moe (2014)

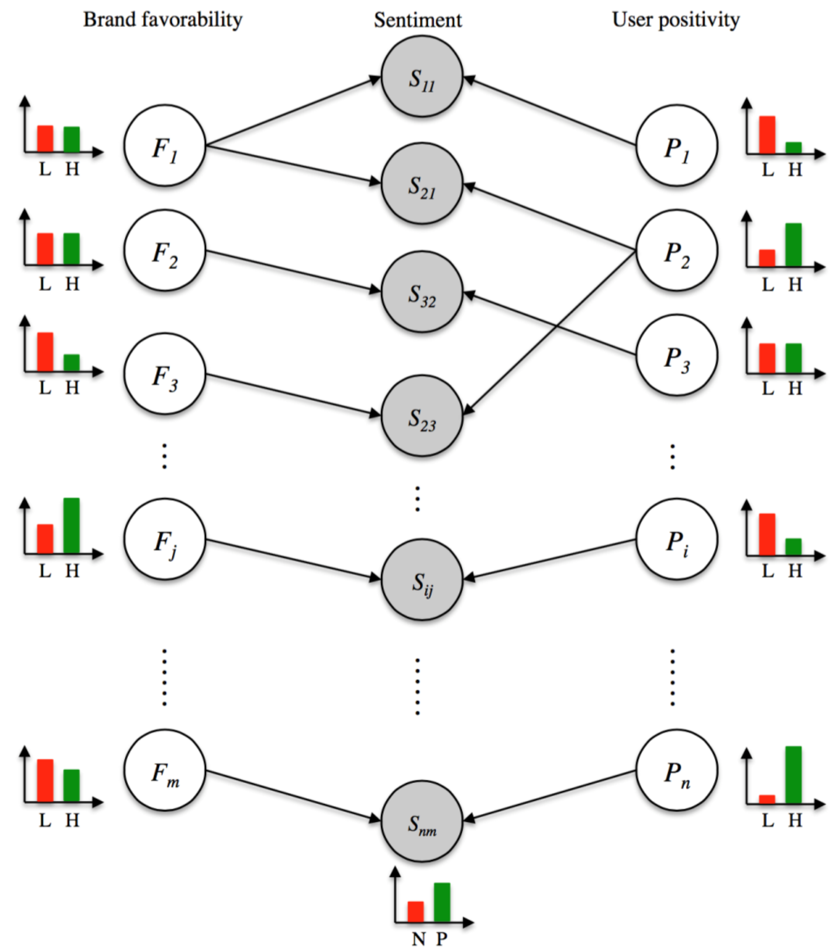
Directional Bias on Social Media Sentiment
(Zhang and Moe 2016)

- Baumgartner and Steenkamp (2001)
  - Response Range (RR) and Midpoint Responding (MPR)
  - Extreme Response Style (ERS)
  - Netacquiescence Response Style (NARS) or Directional Bias
  - Etc.
- This has implications for benchmarking across brands
- Positivity/Negativity is a user trait.



The context of social media

- Repeated observations are obtained from user behavior across brands.
- Users post to differing numbers and unique sets of brand pages.
- Sentiment is observed while brand favorability and user positivity (which both drive sentiment) are latent and estimated.
- For simplicity, assume all variables are binary and has its own probability distribution.



Facebook Data Collection

- Data pertaining to Facebook fan pages related to all brands from English-speaking countries
  - Likes and comments (including sentiment of posted text)*
  - Brand posts
  - Follower metrics (all followers, # followers that like, # followers that comment)

- Dataset description
  - All historical data up to January 1, 2016 (first brand post in our data was in January 2009)
  - 3,355 brand pages from over 20,000, selected based on (1)English language of content and (2) 2015 activity (posts, comments or likes)
  - 273,325,108 users

* Note that the "share" button was launched in late 2011, hence we do not use it in this paper because of lack of data consistency over the entire time period of analysis.

Data Description

- Data cleansing
  - Users who made <5 comments over entire data period
  - Fraudulent activity
    - Users that "liked" posts on >150 different brand pages
    - Users who commented on >100 different brand pages
    - Users who "liked" >90% of posts on a given brand page
    - Duplicate comments (often with URL links to phishing pages)

Number of brands	3,355
Number of brand posts	11,253,623
Number of unique users	169,574,532
Number of user comments	947,550,458
Number of likes	6,681,320,439

Sentiment Coding

- Sentiment identification algorithm applied to the text of each user-posted comment (Zhang et al 2011, Xie et al 2013)
  - Component I: Basic compositional semantic rules
    - Example: If a sentence contains a connective key word like "but", then consider only the sentiment of the "but" clause
  - Component II: Sentiment strength of a phrase
    - Example: "Easy" has a score of 4.1, "best" 5.0, "never" -2.0, and "a bit" 0.03
  - Component III: Special characters (e.g., emoticons, negation words and domain specific words)

- Binary coding of comments
  - Random-forest machine learning model is applied to the above 3 components, resulting in sentiment scores between 0 and 1.
  - Positive if greater than threshold (τ) and non-positive if less than.

- Likes are coded as positive sentiment interactions.

2015 Brand Favorability Scores

Brand Z's Top 20	Favorability score	Range across param settings
Google	0.826	[0.746,0.826]
Microsoft	0.780	[0.753,0.78]
IBM	0.850	[0.841,0.854]
Visa	0.762	[0.75,0.769]
ATT	0.634	[0.594,0.634]
Verizon	0.739	[0.731,0.754]
Coca-Cola	0.782	[0.767,0.789]
McDonald's	0.702	[0.672,0.702]
Facebook	0.877	[0.809,0.877]
Alibaba	0.783	[0.744,0.783]
Amazon.com	0.799	[0.768,0.799]
Wells Fargo	0.841	[0.841,0.888]
GE	0.818	[0.818,0.848]
UPS	0.863	[0.854,0.87]
Disney	0.994	[0.963,0.994]
MasterCard	0.814	[0.732,0.814]
Vodafone UK	0.657	[0.644,0.669]
SAP	0.796	[0.766,0.796]
American Express	0.775	[0.772,0.789]
Wal-Mart	0.769	[0.741,0.769]

Average = 0.793

Brand Z's Bottom 20	Favorability score	Range across param settings
Ford	0.681	[0.646,0.681]
BP	0.720	[0.72,0.765]
Telstra	0.703	[0.674,0.703]
KFC	0.692	[0.689,0.704]
Westpac	0.655	[0.64,0.655]
LinkedIn	0.726	[0.726,0.749]
Santander Bank	0.691	[0.684,0.696]
Woolworths	0.723	[0.723,0.754]
PayPal	0.640	[0.64,0.674]
Chase	0.693	[0.689,0.727]
ALDI USA	0.790	[0.769,0.79]
ING	0.810	[0.79,0.81]
Twitter	0.711	[0.704,0.729]
Nissan	0.788	[0.783,0.81]
Red Bull	0.701	[0.673,0.701]
Bank of America	0.739	[0.695,0.739]
NTT DOCOMO	0.600	[0.6,0.616]
Costco	0.661	[0.63,0.661]
SoftBank	0.633	[0.62,0.643]
Scotiabank	0.687	[0.687,0.705]

Average = 0.702

Validation of Method: Comparison with BrandZ rank

DV = ln (BrandZ Value)

	Estimate	Std. Error	p-value
Intercept	8.253	0.646	<.0001
Favorability score	2.336	0.833	0.006

Adj R^2 = 0.088

	Estimate	Std. Error	p-value
Intercept	10.833	0.493	<.0001
Average Sentiment	-1.046	0.651	0.112

Adj R^2 = 0.019

DV = BrandZ Rank

DV = ln(BrandZ Value)	Estimate	Std. Error	p-value
Intercept	130.940	22.454	<.0001
Favorability score	-114.879	28.954	0.0002

Adj R^2 = 0.163

DV = ln(BrandZ Value)	Estimate	Std. Error	p-value
Intercept	15.544	17.884	0.387
Average Sentiment	36.025	23.639	0.131

Adj R^2 = 0.016

Also considered models with average sentiment, number of likes, number of comments, and number of followers. Adj R^2 is a little better but none are significant predictors.

Implications for Brand Communities

- Small brand communities with limited variance in opinions are more subject to (positive) bias
  - Bias is affected by social media community dynamics
  - Other factors involving brand traits (e.g. industry, product category, size of firm, general awareness/popularity) or activity (e.g., news mentions, social media posting activity) do not affect the bias.
  - Possible echo chamber effects.

- Couldn't positivity bias help the brand through positive word-of-mouth effects?
  - No, average sentiment was not predictive of BrandZ but de-biased brand favorability score was.

Conclusions

- Individuals are subject to various types of behavioral bias on social media.
  - Sentiment is a function the venue and focal topic
  - Social dynamics over time can also systematically bias social media sentiment
  - In a brand community, bias is a function of variance in sentiment and number of followers
- Bias exists on social media, and methods that measure brand favorability need to account for various sources of bias.

Thank you!

Questions?