“It’s Common Sense!”

The Unseen Role of Psychological Theory in Big Social Data Analysis

Candice Lanius
PhD Candidate in Communication & Media
Rensselaer Polytechnic Institute

Albany Medical Center - Journal Club, 19 May 2016
I study what people do with the new found data from social media and information communication technologies (ICTs).

Range of uses:

- Small, personalized data – Quantified Self
- Big, generalized data – Big Social Data
Small, Personalized Data

*The Hidden Anxieties of the Quantified Self Movement*

- “I argue that the use of self-monitoring and tracking technologies can create anxiety around the data capturing process. Tracking technologies create an ordering of people and experiences that discourages moments which are not easily quantified.”

- [https://thesocietypages.org/cyborgology/2015/05/05/the-hidden-anxieties-of-the-quantified-self-movement](https://thesocietypages.org/cyborgology/2015/05/05/the-hidden-anxieties-of-the-quantified-self-movement)
DATA CAPTURE

Loss of rich, subjective complexity
DATA CAPTURE

Constriction ("anxieties") from collecting data
INTERESTING APPLICATIONS OF SMALL DATA

Daily life
- Mood
- Calorie/ Macronutrients
- Sleep
- Movement
- Heart rate

Therapeutic/ Medical

Dr. Deborah Estrin – Small Data Lab – Cornell Tech

“Open mHealth works with clinical experts and app and device developers to make digital health data as useful and actionable as possible.”

- Detect depression and post-partum depression
- Track symptoms and supply interventions

What are the potential consequences of using small data applications as part of the therapeutic process?
BIG SOCIAL DATA ANALYSIS

“…examines large amounts of [social media] data to uncover hidden patterns, correlations and other insights”.
**TWITTER RESEARCH**

**Predict**
- Stock market performance
- Election outcomes
- Flu outbreaks
- Crime
- Cyberbullying
- Protests and revolutions
- Mood/ feelings
- Discriminatory neighborhoods

**Detect**
- Earthquakes
- Traffic-jams
- Allergy sufferers

**Describe**
- Information diffusion
- Public Relations performance

**Predict**
- Donations to charity
- Credible/ False information
- Content recommendations
- User preferences
- User’s location

**Information seeking behaviors**
- Extremism (political)
- Social support for disabilities
EXAMPLE: GOOGLE FLU TRENDS
WHAT IS THE NATURE OF SOCIAL MEDIA DATA?

Taken individually
- Identity construction
- Interpersonal communication
- Brand management
- Advertising
- Activism

Taken in aggregate
- Terministic Screen (Burke)

Results are partially a reflection of the researcher’s agenda, so big social data analysis should include an understanding of rhetoric.
Fundamental changes to research design:

1. Data precede the research questions.

2. Everything is computer mediated.

3. Data is collected from for-profit platforms: not “natural” social interactions.
MY FRAMEWORK: RESEARCH DESIGN AS ARGUMENTS

Data
  / \
Ground  Warrant
     /    /
  Backing  Backing

{Qualifier} Claim
  /    /
    (Rebuttal)
EXAMPLE: “TWITTER MOOD PREDICTS STOCK MARKET”

Twitter (~10 million Tweets) Predicts stock market performance

- Emotional states impact individual decision making (behavioral economics)
- Sentiment analysis of tweets correlated to Dow Jones Industrial Average performance over time
- Collective states may impact systemic decisions
- Mood determined with text analysis (Only certain moods predictive: calm and happy)
  Time series lag of 3 to 4 days is effective
Where are data scientists getting their ground and backing?

Phrased another way:
- How do they decide on research questions?
- How do they generate an explanation for observed patterns?

Common answers:
- Common sense
- Psychology/ Social Psychology
DEBUNKED THEORY: PHRENOLOGY

Correlate physical appearance to behavioral patterns to infer unknown subjects behavior.

19th Century Phrenology
Measure the human skull to determine intellect and personality.
CONCEPT CREEP: Dr. Haslam

Psychology’s technical terms are adapted and applied in new settings without a firm understanding of their clinical significance

- Horizontal - qualitatively new phenomena
- Vertical - capture quantitatively less extreme phenomena

Big social data: cyberbullying, depression, mood states
QUESTIONS FOR DISCUSSION

- What should be the future role of psychologists and other behavioral/social scientists in explaining big social data results?

- What are the limits to applying diagnostic categories outside of therapeutic settings: can the technical or clinical concepts be adapted to large-scale analyses?
EXAMPLE: AttitudeBuzz: Using Social Media Data to Localize Complex Attitudes

Data

Twitter

Qualifier

better than baseline model

Claim

Detect friendly areas for vulnerable populations

Ground

Hateful speech is shared on social media, but linguistic reclamation creates "noise" in dataset

Ground

Differentiate between perjorative and positive speech

Backing

Linguistics Sociology

Backing

Sentiment analysis of geocoded data
Machine learning model
Train against known neighborhoods
EXAMPLE: Social Ties and Check-in Sites

**Data**

- Location Based Social Networks: Gowalla, Brightkite, and Yelp

**Ground**

- Homophily: social bonds form between individuals who share interests

**Backing**

- Social psychology, web science, artificial intelligence research

**Claim**

- Shared checkin locations imply two users are socially connected; Social connections do not indicate shared checkin locations.

**Ground**

- Compare the latent structure of socially connected checkin networks to strangers checkin networks

**Backing**

- Checkin frequencies and social connection for latent structure
- Clustering coefficients statistically compared to underlying network
EXAMPLE: A Tempest in a Teacup? Analyzing Firestorms on Twitter

**Data**
- Twitter (firestorms)

**Ground**
- There are biographical and social consequences to activism

**Back**
- Sociology: survey of high risk activists in Mississippi (1964-1984)

**Qualifier**
- Compared to baseline for one dimension

**Ground**
- Mention networks pre-, during, and post-firestorm

**Back**
- Archived 10% sample of all tweets from time range
- Mention networks as proxy for social ties
- Confidence intervals for difference in global network metrics
- Jaccard index for edges in common

**Claim**
- Firestorms do not change existing social structure