Social Media for Hypothesis Testing

David A. Broniatowski

Department of Engineering Management and Systems Engineering

The George Washington University

Scientific value: Hypothesis testing

GW

- Social media and survey data complement one another, in terms of:
  - Demographics
  - Clock speed
  - Immediacy
  - Analytic rigor
- Social media analysts are on the cusp of developing the research norms and practices that characterize high quality survey research
  - We can test hypotheses rapidly with very large samples
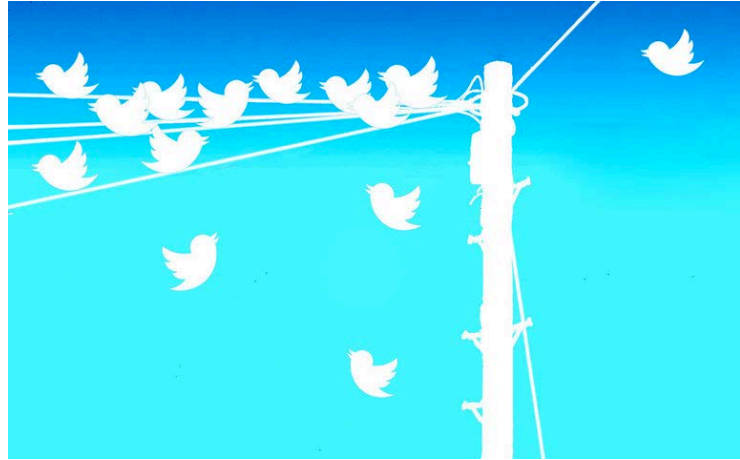
Source:http://media2.govtech.com/images/770*1000/Flickr_Twitter_Telephone.jpg

Dredze, M., Broniatowski, D. A., Smith, M. C., & Hilyard, K. M. (2016). Understanding Vaccine Refusal: Why we need social media now. *American Journal of Preventive Medicine*, *50*(4), 550–552. http://doi.org/10.1016/j.amepre.2015.10.002

Study setting: Disneyland measles outbreak GW

- Began December 2014 at Disneyland in California

- Led to 111 cases in seven states (as well as Canada and Mexico)

- Cases began among unvaccinated people

- Called attention to the issue of herd immunity

- Led to proposals to curtail vaccine refusal through legislative means

Broniatowski, D. A., Hilyard, K. M., & Dredze, M. (2016). Effective vaccine communication during the Disneyland measles outbreak. *Vaccine*. http://doi.org/10.1016/j.vaccine.2016.04.044

Statistics, Stories…. or gist?

GW

- Ongoing debate: Statistics vs. stories ("either-or")
  - Does including a story lead to more effective communications than presenting "just the facts" (i.e., statistical data)?
  - Hesitance to include stories because of concerns of appearing biased
- Fuzzy Trace Theory: Gist and verbatim encoded **in parallel**
  - Verbatim representation (statistical details)
    - "Measles can lead to pneumonia, deafness, lifelong brain damage, and even death, and almost 1/3 of children with measles have to be hospitalized"
  - Gist: Communicates bottom-line meaning
    - "Taking any risk that your child could get the measles and suffer serious complications isn't worth it. Vaccination is the best way to protect your child"
  - Stories are effective because they communicate a gist.
    - Also cue **motivationally relevant moral and social principles**

Reyna, V. F. (2012). Risk perception and communication in vaccination decisions: A fuzzy-trace theory approach. *Vaccine*, *30*(25), 3790–3797.

Analysis of measles media coverage: no response bias

GW

- Coded 4,581 out of a collection of 39,351 outbreak-related articles published from November 2014 to March 2015

- Measured shares on Facebook

- Used M-Turk to categorize article content:

- 1) statistics about viruses or vaccines
  2) "gist", or bottom line meaning
  - Positive or negative summary opinion about endorsing or opposing vaccination

- 3) Other expected covariates based on prior literature

Broniatowski, D. A., Hilyard, K. M., & Dredze, M. (2016). Effective vaccine communication during the disneyland measles outbreak. *Vaccine*. http://doi.org/10.1016/j.vaccine.2016.04.044

What led to article shares:

GW

- Results are consistent with Fuzzy Trace Theory
  - Significant effects of gist and verbatim, but NOT stories
- Stories are effective to the extent that they communicate gist
- Among articles with gists shared at least once (n=257) articles expressing positive opinions about those endorsing vaccination AND those opposing vaccination were 57.8 times more often

Coefficients of logistic regression analysis for whether an article was shared at least once on Facebook (n=4580, df=10)
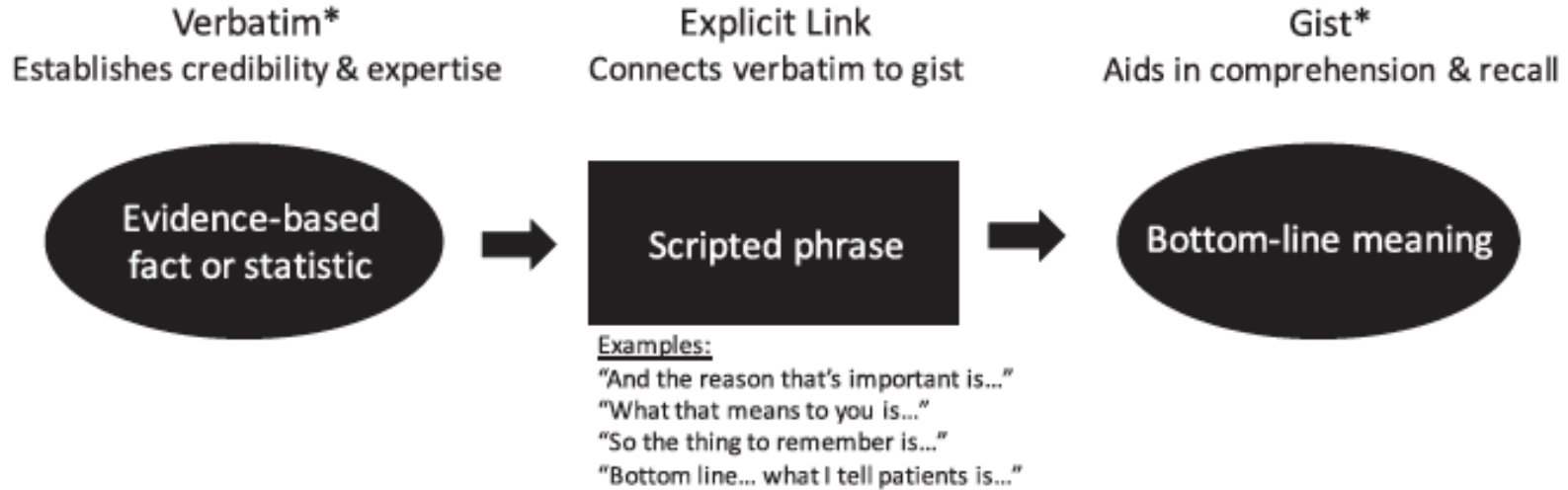
	β	SE β	Z-value	OR
Length	-.000556	.0000893	-6.22*	1.00
Readability	-.000723	.00149	-0.49	1.00
Image	0.59	0.09	6.91*	1.80
Stories	0.34	0.19	1.82	1.41
Statistics	0.29	0.08	3.48*	1.33
Gist	0.82	0.15	5.36*	2.27
Stories x Statistics	0.05	0.22	0.24	1.05
Stories x Gist	0.25	0.32	0.80	1.29
Statistics x Gist	-0.17	0.20	-0.85	0.85
Stories x statistics x Gist	-0.35	0.40	-0.89	0.70
(Intercept)	-1.08	0.12	-8.91*	

Note. * = $p < 0.001$. β = logistic regression coefficient; SE β = standard error of β; OR = Odds Ratio

Broniatowski, D. A., Hilyard, K. M., & Dredze, M. (2016). Effective vaccine communication during the disneyland measles outbreak. *Vaccine*. http://doi.org/10.1016/j.vaccine.2016.04.044

Proposed Gist Communication Framework

Patient-provider communication technique based on Fuzzy-Trace Theory* [10, 11]

Verbatim*
Establishes credibility & expertise

Explicit Link
Connects verbatim to gist

Gist*
Aids in comprehension & recall

Evidence-based fact or statistic → Scripted phrase → Bottom-line meaning

Examples:
"And the reason that's important is…"
"What that means to you is…"
"So the thing to remember is…"
"Bottom line… what I tell patients is…"

Broniatowski, D. A., Hilyard, K. M., & Dredze, M. (2016). Effective vaccine communication during the disneyland measles outbreak. *Vaccine*. http://doi.org/10.1016/j.vaccine.2016.04.044
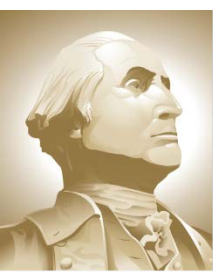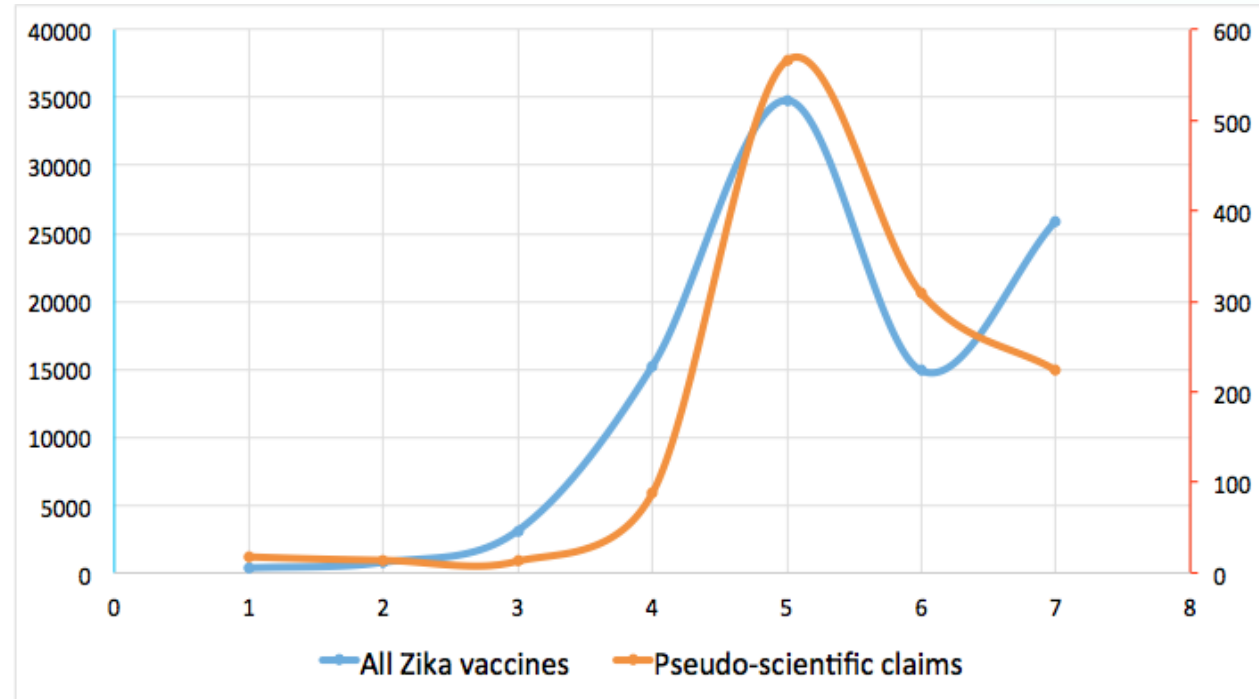
Online Misinformation and Disinformation

James Gathany / CDC / MGN

Dredze, M., Broniatowski, D. A., & Hilyard, K. M. (2016). Zika vaccine misconceptions: A social media analysis. *Vaccine*. http://doi.org/10.1016/j.vaccine.2016.05.008

Tracking Pseudoscientific Claims

- 85% of users previously tweeted about vaccines in 2015

- At least 57% of users previously tweeted an anti-vaccine message

Dredze, M., Broniatowski, D. A., & Hilyard, K. M. (2016). Zika vaccine misconceptions: A social media analysis. *Vaccine*. http://doi.org/10.1016/j.vaccine.2016.05.008

# As if bots weren't bad enough already, now they're anti-vaccine

A dose of the measles-mumps-rubella vaccine. (Damian Dovarganes/AP)

Hypotheses

- Do bots post about vaccines more than the average Twitter user?

- Are bots more likely to be pro-vaccine or anti-vaccine?

- Methods
    - Identified public lists of different types of bot and troll accounts
    - Searched through vaccine stream for their tweets
    - Examined relative proportions of neutral, pro-vaccine, and anti-vaccine tweets
    - Used Bot-o-Meter API to measure the "bot scores" of 9,994 randomly sampled tweets

GW

***=p<0.001, **=p<0.01, *=p<0.05. All results remained significant after controlling for multiple comparisons using Holm-Bonferroni procedure
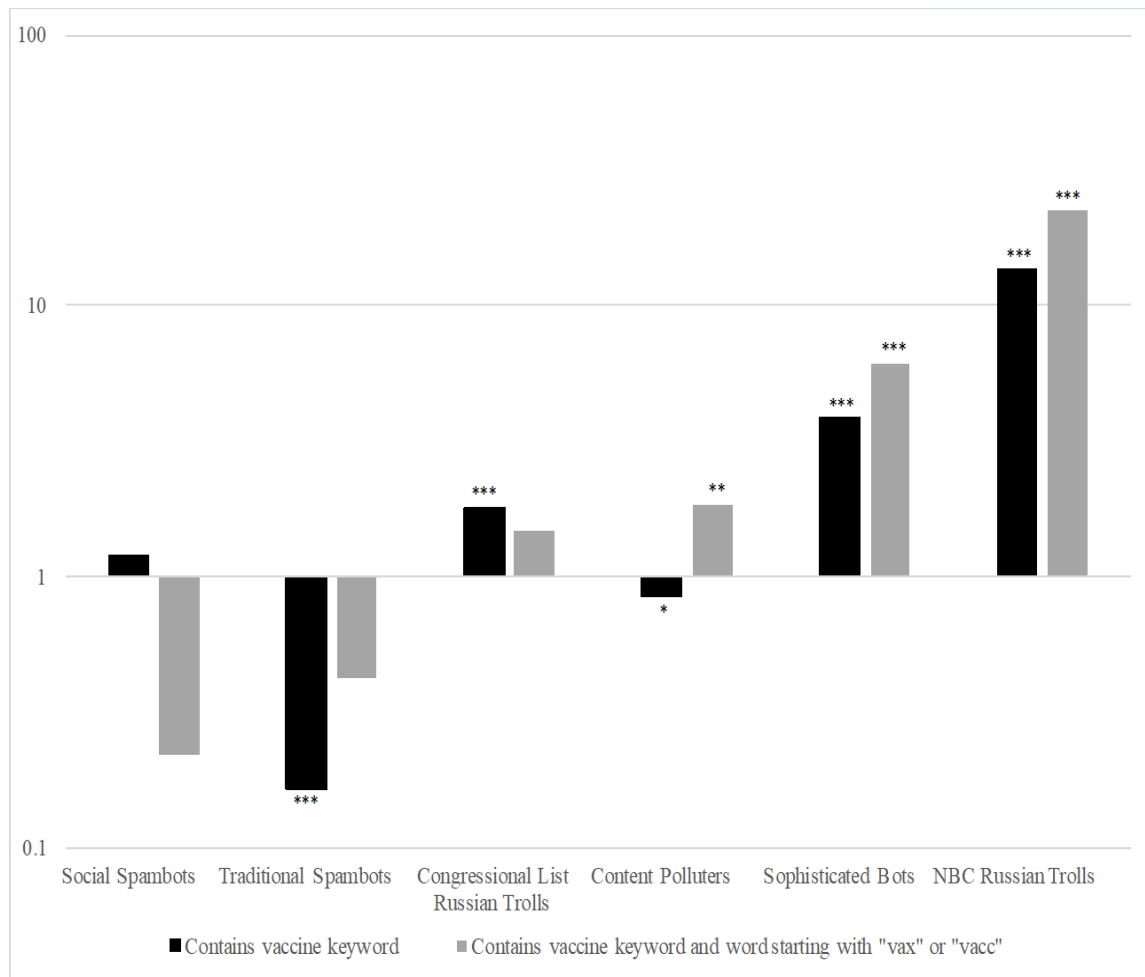
"Unidentified" Accounts and Content Polluters are Most Likely to be Anti-Vaccine

	Polarized (%)	Anti-Vaccine (%)
Assorted Users		
Bot Score < 20%	31	35
Bot Score between 20% and 80%	39***	60***
Bot Score > 80%	26	49*,x
Bot Score unknown	37*,x	62***
Known Bots and Trolls		
NBC Russian Trolls[20]	20*,x	47
Content Polluters[21]	38	60***
Fake Followers[22]	0	N/A
Traditional Spambots[23,24]	3***	0
Social Spambots[23,24]	18**	56*,x
Sophisticated Bots[25]	28	44
Congressional List Russian Trolls[26]	39	48

Conclusions

- Online organized misinformation and/or disinformation campaigns can undermine public health

- In partnership with our collaborators, we are developing new techniques to assess how compelling and influential messages might be
  - Based on empirically validated theory: Fuzzy Trace Theory
  - Complementary to existing survey techniques

- Future directions: Better understanding the drivers of coherent gist communications in online messages
  - How these vary across sociodemographic groups (e.g., different interpretations of emergency use authorization versus standard vaccination)