Understanding Anti-Vaccination Attitudes in Social Media

Tanushree Mitra[1,2]
[1]Georgia Institute of Technology
tmitra3@gatech.edu

Scott Counts[2]
[2]Microsoft Research
counts@microsoft.com

James W. Pennebaker[2,3]
[3]University of Texas at Austin
pennebaker@mail.utexas.edu

Abstract

The anti-vaccination movement threatens public health by reducing the likelihood of disease eradication. With social media's purported role in disseminating anti-vaccine information, it is imperative to understand the drivers of attitudes among participants involved in the vaccination debate on a communication channel critical to the movement: Twitter. Using four years of longitudinal data capturing vaccine discussions on Twitter, we identify users who persistently hold pro and anti attitudes, and those who newly adopt anti attitudes towards vaccination. After gathering each user's entire Twitter timeline, totaling to over 3 million tweets, we explore differences in the individual narratives across the user cohorts. We find that those with long-term anti-vaccination attitudes manifest conspiratorial thinking, mistrust in government, and are resolute and in-group focused in language. New adoptees appear to be predisposed to form anti-vaccination attitudes via similar government distrust and general paranoia, but are more social and less certain than their long-term counterparts. We discuss how this apparent predisposition can interact with social media-fueled events to bring newcomers into the anti-vaccination movement. Given the strong base of conspiratorial thinking underlying anti-vaccination attitudes, we conclude by highlighting the need for alternatives to traditional methods of using authoritative sources such as the government when correcting misleading vaccination claims.

Introduction

Measles, a highly contagious respiratory disease responsible for an estimated 122,000 deaths worldwide each year, was officially eradicated from the United States in 2000. Yet the disease appears to be rebounding. According to the CDC, in 2014 the number of measles cases had reached a 20-year high (CDC 2015). Sadly, many of these cases could have been prevented, as 90% of measles cases in 2014 were in people who were not vaccinated or whose vaccination status was unknown. One reason for this rebound is that concerns about vaccine side effects have taken precedence over the dangers of potentially deadly vaccine-preventable diseases and a vaccination culture promoting anti-vaccination has emerged (Kata 2010). This persistent vaccine criticism movement has spread rapidly through social media, a channel often used to disseminate medical information without verification by the expert medical community (Keelan et al. 2010).

Given the increasing reliance on online media for accurate health information and the general growth of social media sites, the attitudes of anti-vaccination advocates risk becoming a global phenomenon that could impact immunization behavior at significant scale (Kata 2010). In fact a controlled study showed that parents opting to exempt children from vaccination are more likely to have received the information online compared to those vaccinating their kids (Salmon et al. 2005). These parents benefit from "herd immunity" in which eradication is achieved by immunizing a critical proportion of the population. However, as internet-fueled misbeliefs drive people to opt out of vaccination, herd immunity is weakened, increasing the chances of a disease outbreak. Thus it is important to understand the underlying characteristics of individuals with anti-vaccination attitudes. What drives people to develop and perpetuate the anti-vaccination movement?

In this paper we explore this question by examining individuals' overt expressions towards vaccination in a social media platform extensively used for vaccine discussions: Twitter. By using four years of longitudinal data capturing vaccination discussions on Twitter, we identify three sets of key individuals: users who are persistently pro vaccine, those who are persistently anti vaccine and users who newly join the anti-vaccination cohort following an event symbolic to the vaccine controversy. Long-term anti-vaccination advocates play an important role in preventing eradication because they sustain weakness in herd immunity, and thus it is crucial to understand them and their motivations. Examining new anti-vaccination proponents allows us to understand the type of person that would adopt such a stance despite strong recommendations to the contrary from authoritative organizations like the CDC. After fetching each cohort's entire timeline of tweets, totaling to more than 3 million tweets, we compare and contrast their linguistic styles, topics of interest, social characteristics and underlying cognitive dimensions, all with an eye to uncovering the drivers of such extreme attitudes against a social good.

We find that people holding persistent anti-vaccination attitudes use more direct language and have higher expressions of anger compared to their pro counterparts. They

Pro vaccination Tweets	Anti Vaccination Tweets
Measles Outbreak in U.K. Linked to Poor MMR Vaccine Uptake After Autism Scare [link]	Would rather have measles in my kid than lifetime seizures and autism, VIT A for a week, big deal #CDCwhistleblower
#Vaccines are not a scientific controversy. They work.	The vaccines have increased autism, peanut allergies & childhood cancers.

Table 1: Example tweets labeled as pro and anti-vaccination by three master Turkers.

also show general conspiracy thinking and mistrust in the government, suggesting characteristics of paranoia. Adopters of anti-vaccine attitudes show similar conspiratorial ideation and suspicion toward government even before they start expressing anti-vaccine attitudes. This suggests that the new adoptees are already predisposed to form anti-vaccine attitudes. Moreover, long-term anti-vaccine advocates exhibit higher sense of group solidarity. Individuals in such close-knit groups typically end up adhering to their extreme positions, often fueling beliefs in false conspiracies and are particularly resistant to correction (Sunstein and Vermeule 2009). These findings suggest that health officials attempting to simply correct conspiracy fuelled false claims might be counterproductive. Thus newer methods to counter the harmful consequences of anti-vaccination beliefs are needed.

Broadly we hope this work contributes to two bodies of research: computer mediated communication (CMC) research on contentious topics (Kata 2010; Keelan et al. 2010; Salmon et al. 2005), and psychology research on maintaining and adopting an attitude (Kristiansen and Zanna 1988; Rokeach 1968; Savion 2012). With respect to CMC research, examining the naturally occurring self-expressions of these cohorts highlight the linguistic style, social media characteristics, and topics of interest driving people with health behavior beliefs that are antithetical to the health of society as a whole. With regards to attitude psychology, while there are numerous lab-based studies listing various factors driving resistant attitudes, our study is one of the first to provide large-scale empirical evidence of factors behind resistant attitudes towards vaccination.

Related Work

We provide a brief overview of two broad areas of research relevant to our study: attitude measurement and text analytic approaches to study user traits.

Measuring Attitude

The concept of attitude has long been central to social psychology research. However, the definition of attitude has changed over the years resulting in no single universally accepted definition (Schwarz 2007). For the purposes of this study, we use the evaluative definition of attitude from the seminal work of Eagly and Chaiken (1993) – "attitude is expressed by evaluating a particular entity with some degree of favor or disfavor". By examining the social media posts from users we determine whether the expressed attitude towards vaccination is for or against it.

Despite a growing body of work on attitude, a major concern in its study has been the problem of accurate measurement. Attitude studies based on questionnaires and self-reports are highly context dependent and results can vary with changes in question wording, format or order (Schuman and Presser 1981). Participants can also conform to the demands of the questionnaire by creating superficial expressions of attitude (Abelson 1988). Even newer methods of implicit measures of attitude (Dovidio and Fazio 1992) have shown sensitivity towards context, raising questions about its effectiveness over self-reported measures. However, the rapid growth of text-based social media has opened new opportunities to study attitudes unobtrusively, as they naturally unfold in large populations and over long time periods. For example, population attitudes extracted from tweet sentiments has been shown to correlate with traditional polling data (O'Connor et al. 2010). Machine learning techniques on textual data have accurately predicted sentence level attitudes in online discussions (Hassan, Qazvinian, and Radev 2010). Drawing on the success of studying attitudes from online textual data, we built a classifier to determine positive and negative attitudes towards vaccination.

Analyzing text to infer individual characteristics

Attitudes and language are intimately related (Eiser 1975). For decades social scientists have demonstrated that individuals often adopt language consistent with their attitudes (Eiser and Ross 1977). Hence a growing number of studies have used content-analytic approaches to assess individual differences and personality characteristics. A popular approach is simply to count and categorize the words that people speak. A validated tool for such an approach is LIWC, Linguistic Inquiry and Word Count (Tausczik and Pennebaker 2010). Researchers have used LIWC for a number of psychological measurement tasks, such as deciphering psychological states of presidential candidates from their spontaneous speech samples (Slatcher et al. 2007), identifying true and false stories by analyzing linguistic styles (Newman et al. 2003), examining differences between depressed and non-depressed individuals (Rude, Gortner, and Pennebaker 2004), and contrasting pro-anorexic and pro-recovery people (De Choudhury 2015).

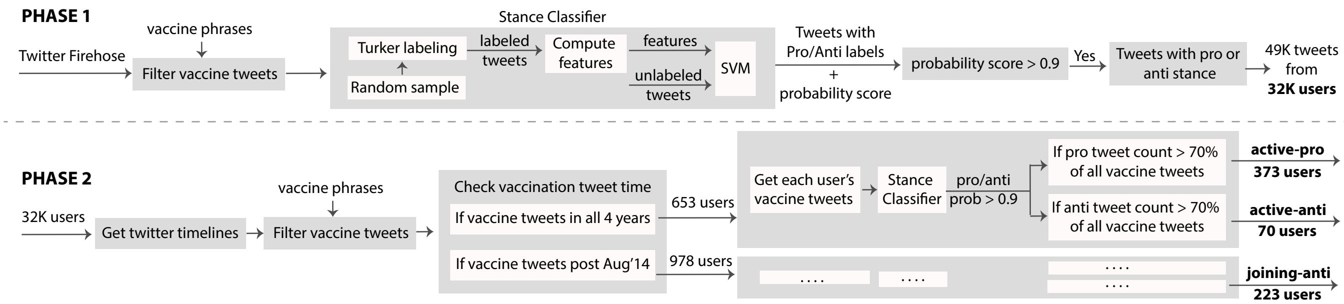

Figure 1: Steps to identify user cohorts. Ellipses (...) shown at the bottom panel of phase 2 correspond to an equivalent module from the top panel.

Another popular text analytic technique among social psychologists is the Meaning Extraction Method (MEM) (Chung and Pennebaker 2008). MEM can reliably infer the dimensions along which people think about themselves or particular issues. Scientists have used MEM to capture psychological dimensions of self expressions in personal narratives (Chung and Pennebaker 2008), measure people's basic values underlying their attitude (Ryan L. Boyd et al. 2015) and capture dimensions of self-expressions in social media posts, such as Facebook status updates (Kramer and Chung 2011). Following in their footsteps, our work uses MEM to capture differences in the dimensions of thought across groups of individuals with differing attitudes towards vaccination. We complement this with LIWC's linguistic analysis to examine differences in expression styles. We return to the details of MEM later.

Data Collection

Our data collection involves two main phases. Figure 1 outlines the main steps.

Phase1: Collecting Pro & Anti-Vaccination Tweets

To fetch vaccine specific posts, we first did a manual examination of 1000 Twitter posts and 50 news reports on vaccination to identify search terms and phrases relevant to the vaccination debate. Based on a snowball sampling approach, we used the initial set of search terms to extract a tweet sample from the Twitter Firehose[2] stream between January 1 and 5, 2012. Two researchers then manually inspected the sample to identify any co-occurring terms missed and to remove spurious terms that might be returning tweets too broad to classify on either side of the issue. The final set had 5 phrases: 'vaccination+autism', 'vaccine+autism', 'mmr+vaccination', 'measles+autism' and 'mmr+vaccine'. Using these phrases we fetched tweets spanning four calendar years – January 1, 2012 to June 30, 2015, totaling to 315,240 tweets generated by 144,817 unique users. Our next task was to build a classifier to identify pro and anti stance of the collected vaccine tweets.

Classify Vaccination Stance – Stance Classifier

Our classifier was based on a two-step labeling process: (1) gathering human annotations on a sample, and (2) leverage the labeled data to annotate the remaining collection of tweets. For the human annotation task, we randomly sampled a batch of 2000 posts from each year (a total of 8000 posts) and recruited 3 independent Amazon Mechanical Turk master workers residing in the United States to identify whether the post content is for or against vaccination. Workers were presented with definitions of pro and anti vaccination content along with example tweets to train them in the labeling task. We retained posts where all three workers agreed on the post being either pro or anti. We show a sample in Table 1.

Based on a qualitative examination of the frequently occurring unigrams, bigrams, trigrams and hashtags, we found that trigrams and hashtags were prominent cues of a tweet's stance towards vaccination. For example, phrases like *vaccines causing encephalitis*, *already #vaccineinjured like, #VaccineInducedAutism* are indicative of the tweet being against vaccination, while *shut-up #antivaxxers #science*, *push mmr vaccine*, *still deny vaccinations* signal pro-vaccination stance. Using trigrams and hashtags as features, we built a supervised learning classifier by training a support vector machine (SVM) under 10-fold cross validation. We refer to this as our "stance classifier". We retained only tweets for which the classifier predicted stance with high probability (greater than 0.9). We purposely choose a high threshold to maintain a high level of precision. The prediction accuracy of our classifier was 84.7%. The resulting dataset had 49,354 tweets, classified with attitude stance and posted by 32,282 unique users. In the next phase, we examine this user set to identify our user cohorts.

Phase2: Identifying User Cohorts

The goal of this phase is to segregate three principle actors – long term advocates of pro and anti vaccination attitude and users newly adopting anti-vaccination attitude. Our

[2] The Twitter Firehose stream is a dataset of all public posts from Twitter made available to us through an agreement with Twitter.

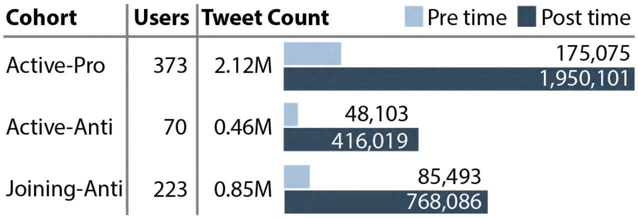

Figure 2: User cohort statistics

criterion for selecting long-term advocates is to identify users who consistently post tweets expressing pro or anti attitudes towards vaccination in all four years (2012 to 2015). To find users who newly joined the anti-vaccination cohort, we needed to fix on a time period before which users did not mention about vaccination but actively tweeted anti-vaccination posts after that time. We selected August 15, 2014 as our cutoff time. This time was marked by a symbolic event pertaining to the vaccine debate: A false claim reported that a CDC whistleblower exposed CDC data linking autism among African-American boys following MMR vaccination[3]. The event was widely tweeted under the #CDCwhistleblower hashtag and was a topic of active discussion among users interested in the vaccine debate.

Collecting Historical Posts from Users

Recall that 32,282 unique users posted our filtered collection of vaccination tweets. We collect their entire Twitter timeline tracing back from June 30, 2015 (time when our data collection started). Using the same set of vaccine related search terms as in phase 1, we search each user's Twitter timeline to find their history of vaccine posts. Next, we pass their vaccine specific tweets through our stance classifier, retaining only those for which the classifier is able to determine pro or anti stance with high probability (>= 0.9). Finally we classify users as holding pro (anti) attitude if more than 70% of their vaccination tweets are classified as pro (anti). Users who were consistently classified as pro or anti across all four years are the long-term advocates. Hereafter we refer to them as *active-pro* and *active-anti.* Users who were classified as holding an anti-vaccine attitude after August 15, 2014 are adoptees of the anti-vaccination attitude (the *joining-anti* cohort). We intentionally did not create a *joining-pro* cohort. While interesting, our focus is on understanding the anti-vaccination users, for whom the *active-pro* cohort can provide contrast as necessary.

Our cohort selection process takes a very conservative approach, resulting in considerable shrinkage of our cohort pool. We purposely opt for this strategy to maintain high precision and to have high confidence that the users are indeed holding these extreme attitudes. Our final cohort comprises 373 *active-pro*, 70 *active-anti,* 223 *joining-anti* users with each group contributing about 2.12M, 0.46M and 0.85M tweets respectively.

Method

What can we learn about people with pro or anti attitudes towards vaccination from their digital footprints left in Twitter? Recall that people's natural language expression can convey personalities and cognitive processes (Cohn, Mehl, and Pennebaker 2004). Hence we adopt a lexical approach to compare and contrast our cohorts. Lexical approaches have been widely used by psychologists to capture personality traits (Allport and Odbert 1936), cognitive processes (Slobin 1996) and more recently to analyze open-ended self-descriptions (Chung and Pennebaker 2008). The basic premise of a lexical approach is that the important ways in which individuals differ are represented by words. Using lexical approaches we can understand *how* people talk while mentioning the vaccination issue and *what* topics they talk about? To investigate the *what* aspect we inductively extract meaningful themes from users' natural language expressions using the Meaning Extraction Method (Chung and Pennebaker 2008). To examine the *how* aspect we use the LIWC program for comparing usage rates of function and content words in tweets from individuals. We complement our lexical quantifications with social media specific features such as user's follower to following ratio and tweet volume.

Meaning Extraction Method (MEM)

MEM is a topic modeling approach to extract dimensions along which users express themselves. For example, an individual with 'vaccination' as an important part of her cognitive self is likely to engage in thoughts related to vaccination or even other related health topics (Cantor 1990; Markus 1977). These cognitive dimensions will naturally be reflected in the words that she uses in her everyday social media posts. MEM uses a factor analytic approach to form clusters of co-occurring words (Chung and Pennebaker 2008). Specifically, it performs a principal component analysis on the matrix of words used by social media post authors to find how they naturally co-occur. These word co-occurrences identify linguistic dimensions that represent psychologically meaningful themes. Contrary to other topic modeling approaches, the MEM extracted dimensions can be examined at the level of an individual, allowing us to make meaningful inferences in the context of a user. Being a word-count based approach, MEM is highly interpretable. Further, MEM has been shown to capture psychological dimensions of self-expressions in personal narratives (Chung and Pennebaker 2008) and in social media posts (Kramer and Chung 2011).

[3] http://www.snopes.com/medical/disease/cdcwhistleblower.asp

Factors	Words
1. Evil Govt.	war, government, arrest, military, iran, terrorist, romney, drone, terror, spy, cia, bomb
2. Autism vs. Affect	hate*, feel*, research, said*, damn*, suck*, national, autism, mean*, genetic
3. Chronic Health	risk, improve, heart, surgery, alcohol, chronic, brain, longterm, patient, stroke
4. Govt. Vote	voter, republican, candidate, conservative, budget, romney, senate, democratic
5. Organic Food	food, organic, product, farm, chemical, healthy, pesticide, genetically, gmo, fat
6. Family vs. Nuance	family, interest*, baby, wrong*, kid, toddler, mother, argument*, evidence*, belief*
7. Technology	apple, space, nasa, galaxy, smartphone android, tablet, iphone, rocket, google
8. Vaccine & Diseases	outbreak, measles, hepatitis, hiv, malaria, vaccine, pandemic, infection, death, polio, tb

Table 2: Themes emerging from the tweets of *active-pro* and *active-anti* users. Representative words with the highest factor loadings are shown per theme. Words with negative factor loadings appear with '*'.

	Between Groups		Within Time	
Factors	Mean Diff	p-val	Mean Diff	p-val
Evil Government	*A > P*	$<10^{-15}$		ns
Autism vs. Affect		ns	Post > Pre	$<10^{-15}$
Chronic Health	*P > A*	$<10^{-15}$		ns
Govt. Vote		ns		ns
Organic Food	*A > P*	0.042	Post > Pre	0.002
Family vs. Nuance	*A > P*	$<10^{-15}$		ns
Technology	*P > A*	0.012		ns
Vaccine & Diseases		ns	Post > Pre	0.001

Table 3: Theme comparisons between *active-pro* (P) and *active-anti (A)* groups. ■ marks themes where A > P and ■ corresponds to P > A. Statistically significant differences are based on an independent sample t-test. 'ns' denotes non-significant results.

Results

Understanding Long-Term Advocates

We want to identify the characteristics of individuals holding persistent attitudes towards vaccination, our *active-pro* and *active-anti* user cohorts.

What topics are relevant to them? To know *what* topics the two cohorts talk about, we passed their combined twitter feeds (2.58M in total) through the Mean Extraction helper software (Boyd 2015) which automates the MEM approach. We identified 8 factors that explain 63.4% of the variability in the user data (Table 2)[4]. Considering the generative nature of language, 63.4% of variance is fairly high (Chung and Pennebaker 2008). Key factors included the factor labeled 'Evil Government', with terms like *war, government, arrest, military, iran, terrorist*. The 'Chronic Health' factor comprised terms referring to persistent health conditions and its associated risks and treatment. For the 'Autism vs. Affect' factor, words describing emotion loaded negatively, while words referring to autism loaded positively, suggesting that this factor is referring to autism in an objective sense, devoid of much emotion. Another important factor that surfaced is 'Vaccine & Disease', with words referring to disease outbreaks (*outbreak, measles, hepatitis, pandemic, hiv*) and vaccination (*vaccine, immunization, trial*). It is worth noting that the emergence of this last factor helps validate the MEM approach, as we would expect long-term advocates to be quite involved in discussions of vaccines and diseases. Having derived the thematic factors, our next task is to examine the differences and similarities across these dimensions. Following between and within subjects experimental design approach, we perform two sets of comparisons: between groups and within time (see Table 3).

To start, we compare the two long-term active user cohorts across the 8 factors (Table 3, Between Group). We find that the *active-anti* group uses more words referring to 'Evil Government', 'Organic Food' and 'Family vs. Nuance. Given the loadings of the specific words on the 'Family vs. Nuance' factor, the *active-anti* cohort appears to be relatively focused on the family aspects of vaccination but in absolute terms, without reflection on beliefs or evidence. The *active-pro* cohort uses 'Technology' and 'Chronic Health' topics more, the latter implying a general interest in healthy living. The differences were not statistically significant across the remaining factors – 'Autism vs. Affect' and 'Government Vote'.

How do the themes differ before and after a user's first reference to vaccination? This analysis helps us capture any significant difference in a user's behavior across these two important phases. To divide a user's twitter timeline into the temporal phases, we locate the timestamp of her first vaccination tweet. Next, we group all tweet content into two respective buckets – *pre* posts, comprising of tweets mentioned before user's first vaccine tweet and *post* content groups tweets after this timestamp (refer Figure 2 for count statistics). Recall that our MEM method had already identified the underlying themes in the entire tweet corpus of the two cohorts. By performing an independent sample t-test across the *pre* and *post* groups, we are able to distinguish the themes across time (Table 3, Within Time). We find that users refer to 'Vaccine & Diseases', 'Autism vs. Effect' and 'Organic Food' more in their *post* time tweets compared to *pre* tweet. This suggests a general alertness towards healthy food consumption and the safety

[4] Similar to other topic modeling methods, researchers have some degree of flexibility to determine the number of themes. For MEM, theme interpretability is the main determining factor.

Social & Identity		**Cog thinking & Tentative**	
Ingroup	A > P*	Cause	P > A*
Emotion		Insight	P > A****
Anxiety	P > A*	Tentative	P > A****
Anger	A > P****	Exclusions	P > A****
Linguistic Style		Interrogatives	P > A****
Assent	P > A****	Conjunction	P > A****
Non-fluencies	P > A*	Certainty	A > P*
		Personal Concerns	
		Work	P > A*
Time Orientation		Death	A > P****
Present Focus	P > A**	**Biological States**	
Engagement		Health	P > A****
Vaccine tweet prop		Sexual	P > A**
Social Media Characteristics			
Attention Status	P > A*	Message Reach	A > P*
Directed Communication		Information content index	

Table 4: Comparing the *active-pro* (P) and *active-anti* (A) cohorts. Statistically significant LIWC results based on independent sample t-tests are listed. Social media characteristics are compared using Wilcoxon rank sum test. *p*-values are shown after BH correction to control familywise error-rate: * p ≤ .05; ** p ≤ 0.01; ***p ≤ .001; **** p ≤ .0001.

of genetically modified food and pesticides following their first online expression on vaccines. They also start talking about autism in a more objective manner, mostly sharing information about the absence or presence of link between vaccine and autism.

How do they present themselves? To investigate this, we turn to our LIWC results in Table 4. Again, we list only results with statistically significant differences across cohorts, and call out a few results of note in the text. For instance, the *active-anti* cohort uses significantly more ingroup language, indicating their effort to invoke and maintain social connections and group solidarity. Emotional differences were noteworthy with the *active-pro* cohort exhibiting greater anxiety and the *active-anti* cohort greater anger. They also exhibited extreme negative concerns through their higher rates of *death* related words. On the contrary, *active-pro* users focused more on the *present*, and had concerns about work and health (higher *health* and *sexual* category words). They also showed higher cognitive processing through their greater usage of *causal* and *insight words*, indicated an informal style by using more *assents* and *non-fluencies* and displayed lower assuredness through increased use of *tentative*, *exclusions*, *interrogatives* and *conjunctions*. In contrast, *active-anti* users were more certain (higher rates of *certain* words) and more direct in their language (fewer *assents* and *non-fluencies*). They also exhibited a drop in immediacy (less present focus), suggesting possible emotional distancing, an often used coping mechanism to deal with acutely upsetting events (Holman and Silver 1998).

Do their social media characteristics differ? Finally we examined the following social media characteristics - 1). *Attention status ratio*, is the ratio of followers (those who pay attention to the user) to following (those the user pays attention to) (Hutto, Yardi, and Gilbert 2013), (2). *Message reach*, the ratio of retweets and favorites to the total number of tweets. (3). *Amount of directed communication* is the ratio of tweets with "@" mentions to total tweet count. (4). *Informational content index* is the ratio of tweets containing a URL. As the distribution of social network based measures deviates from normality, we compare these measures across the cohorts using the Wilcoxon signed rank test. We find that *active-pro* had greater *attention-status ratio* while *active-anti* users had more *message reach*. The extent of directed communication and informational content index was not statistically different. Additionally, to test for differences in the degree of engagement in the vaccine issue, we compute engagement as a proportion of tweets related to vaccination in the entire user timeline. We did not find any statistically significant difference, suggesting that both the long-term advocates were equally vested in forwarding their opinions towards vaccination.

Understanding Anti-Vaccination Joiners

To understand the characteristics of people who join the anti-vaccination movement, we compare recent adoptees (*joining-anti*) with those already perpetuating the movement (*active-anti*).

What topics are relevant to them? To start, we pass the entire Twitter timelines of *active-anti* and *joining-anti* users (1.32M posts in total) through the MEM analysis as before. We found that 10 factors, explaining 45.4% of the variance emerged from our dataset (see Table 5). We label Factor 1 'War News' because of its reference to terror and war (*terrorist, rebel, Syria*). The factor, 'Secret Government' contains terms referring to government's attempt to conceal information (*secret, cia, underground, caught, alert*). A parallel factor emerging from our analysis is 'Government Conspiracy' with terms alluding to covert schemes (*bilderberg, imf, conspiracy, politics*). Here the term Bilderberg refers to the Bilderberg group, which has often been accused of conspiracies (see ("Bilderberg Group" 2015)). Another similar theme is the 'Vaccine Fraud' factor with terms referring to disbelief in vaccination and government's attempt to hide the adverse consequences of vaccination (*coverup, omit, data, cdcwhistleblower*). In particular, *cdcwhistleblower* was a frequently used hashtag by people from the anti-vaccination camp to refer to the popular hoax story linking autism and vaccines ("Snopes.com" 2015).

Comparing factors across the *active-anti* and *joining-anti* user groups shows that *joining-anti* group refers more to 'Vaccine Fraud', whereas *active-anti* are more likely to

Factors	Words
1. War News	terror, war, terrorist, foreign, rebel, regime, west, turkish, russian, syria, qaeda, humanitarian
2. Secret Govt.	secret , illuminati, homeland, underground, cia, camera, fema, wtc, ufo, agent, sight, caught
3. Informal Speech	lol, feel, em, tomorrow, room, outside, gonna, luck, coffee, cat, bet, suck, easy, annoy, dumb
4. Govt. Vote	tax, voter, romney, obama, president, medicare, democrat, capital, republican, donor, gop, poll
5. Vaccine Fraud	cdcwhistleblower, vaccinate, mmr, fraud, data, investigate, measles, thomspon, coverup, harm
6. Chronic Health	lung, obesity, detect, muscle, clinical, regulation, diabetes, implant, hip, acid, brain, addiction
7. Organic Food	chemical, farmer, food, crop, gmo, usda, organic environmental, genetically, pesticide, monsanto
8. Government Conspiracy	imf, bilderberg, dictatorship, intimidate,eugenic, embassy, laden, conspiracy, politics, infowar
9. War & Terror	dispute, catastrophe, osama, felony, nuke, crisis, helicopter, missile, militia, enforcement, federal
10. Religious Extremism	hamas, Nazis, german, muslim, union, evil, nigeria, accept, warn, refugee, christian, silence

Table 5: Themes emerging from the social media posts of *active-anti* and *joining-anti* users. Representative words with the highest factor loadings are shown per theme.

refer to 'War News', 'Secretive Government', 'Government Vote', 'Chronic Health' and 'Government Conspiracy' (see Table 6). This provides additional evidence that those joining the anti-vaccination camp are specifically concerned about vaccination and a potential vaccine fraud, while the long-term anti-vaccination supporters have a broader agenda of government distrust. Comparisons over time show that *anti* users refer more to 'Vaccine Fraud' and 'Chronic Health' *post* their first vaccine tweet.

How do they present themselves? As before, to understand how these two groups participate in the vaccine discussion, we turn to comparisons across the LIWC measures (Table 7), calling out a few results of note. The *active-anti* users demonstrated higher cognitive complexity by using complex sentences (based on high preposition usage) and exhibited concrete thinking through their increased use of concrete nouns (*articles*). In contrast, *joining-anti* users signaled lack of definitiveness through higher usage of *interrogation, discrepancy, negation, exclusions* and *conjunctions* words. Consistent with their cognitive concreteness, *active-anti* users also showed emotional distancing by their lower usage of *positive emotion* words compared to *joining-anti*. Also in line with their overly concrete expressions, *active-anti* group had higher personal concern for *money* and *work*. *Joining-anti* had higher rates of *leisure* words, which is consistent with their higher *positive emotion* word usage. They were also more *present* focused and exhibited increased social orientation (higher *social process* and increased *2nd person pronoun* usage).

	Between Groups		Within Time	
Factors	Mean Diff	p-val	Mean Diff	p-val
War News	*A > J*	$<10^{-15}$		ns
Secretive Govt.	*A > J*	$<10^{-15}$		ns
Informal Speech		ns		ns
Govt. Vote	*A > J*	$<10^{-15}$		ns
Vaccine Fraud	*J > A*	$<10^{-15}$	Post >Pre	$<10^{-15}$
Chronic Health	*A > J*	$<10^{-15}$	Post >Pre	0.002
Organic Food		ns		ns
Govt. Conspiracy	*A > J*	0.005		ns
War & Terror		ns		ns
Religious Extremism		ns		ns

Table 6: Theme comparisons between the active-anti (A) and joining-anti (J) groups. Statistically significant differences are based on an independent sample t-test. ■ marks themes where A > J. ■ correspond to themes where J > A.

Do their social media characteristics differ? We find that *active-anti* had greater *attention-status ratio* compared to *joining-anti*. The extent of directed communication, informational content and message reach was not statistically different between the two user cohorts.

Taken together, these comparisons suggest that those joining the anti-vaccination cohort are more social and less definitive – indicators of people who might join a cause or a group. Long-term anti-vaccination supporters who are concrete and complex in thought had higher attention status ratio – indicators of people who could perpetuate a cause. Those joining also posted relatively more content about vaccination (higher *engagement*), suggesting that for them this was, at least initially, a specific issue of interest, while for long-term anti-vaccination advocates, vaccination appears to be one in a number of government conspiracy issues of interest. This general conspiracy thinking and cognitive mindset thus appears to provide the perfect net to catch people specifically concerned with vaccination.

Discussion

Anti-vaccine advocates manifest conspiracy thinking

Themes emerging from the twitter history of anti-vaccine advocates refer to *government conspiracy*, deliberate *vaccine frauds*, accusations of cover-ups by regulatory bodies ('Secretive Government') and concerns over terror wars ('War News') (see Table 5). In fact they had significantly higher mentions of *vaccine fraud* and *chronic health* concerns after their first mention of vaccination. This suggests an important characteristic of people holding unfavorable attitude towards vaccination: an inclination towards conspiracy thinking – a way of interpreting the world where conspiracy plays a dominant role (Zonis and Joseph 1994). It is worth noting that these conspiracy related topics surface as prominent themes amidst the millions of tweets expressed over multiple years. This suggests that conspira-

Social & Identity		Cog think & Tentative	
2nd person pronoun	J > A****		
3rd person pronoun	J > A**	Discrepancy	J > A*
Imperson Pronoun	J > A****	Negation	J > A****
Social Process	J > A****	Exclusions	J > A**
Affect & Emotional Distance		Interrogation	J > A****
Positive Emotion	J > A*	Conjunction	J > A*
Articles	A > J****	Insight	J > A****
Prepositions	A > J**		
Time Orientation		**Personal Concerns**	
Present Focus	J > A****	Leisure	J > A****
Engagement		Money	A > J*
Vaccine tweet prop	J > A**	Work	A > J*
Social Media Characteristics			
Atten Status ratio	A > J**	Message Reach	
Directed Communication		Information content index	

Table 7: Comparing the *active-anti* (A) and *joining-anti* (J) cohorts using LIWC measures. *p*-values after BH correction to control familywise error-rates are shown.

torial ideation is not only driving their anti-vaccination beliefs but that the very notion of conspiracy has a strong hold on their way of reasoning about events in the world.

> "Freedom of information request reveals major government **vaccine conspiracy** gaia health #autism #aspie"
>
> "Chemtrails/Death-Dumps! **Secret Govt Operation** | USAHM Conspiracy News [link]"
>
> "9/11 Blueprint for Truth – The Most Compelling Presentation Proving the 9/11 **Conspiracy** … [link]"

The above example tweets from our *anti* users also align with research showing people's consistency with conspiratorial ideas: someone who believes in one conspiracy theory, tends to believe others as well (Swami et al. 2011). This consistency is an artifact of people's tendency to maintain a coherent system of attitudes so as to strike internal and psychological consistency (Bem 1970). For individuals with anti-vaccine attitudes, a paranoid world of conspiracy theories, secret, sinister organizations and manipulative government bodies causing harm are all part of their coherent system of beliefs.

Resoluteness towards anti-vaccination stance

Our results present evidence of the strength and persistence of the underlying attitude conviction of long-term anti-vaccine advocates. They showed relatively higher usage of concrete nouns, an indication of definitive expression (Table 4). They even used more certainty terms compared to their pro counterparts. This aligns with research examining dimensions of attitude strength and resistance towards change (Pomerantz, Chaiken, and Tordesillas 1995): a primary dimension of attitude strength is the degree of expressed certainty. In contrast, *active-pro* advocates used more indirect language.

The emerging MEM themes of conspiratorial worldview also support *active-anti* cohort's persistence towards anti-vaccine attitudes. Social psychologists have argued that conspiratorial thinking serves as a "cognitive shortcut" that can be used to simplify and explain larger, more complex effects. This sense making function can lead people to conspiratorial beliefs despite little evidence to warrant such beliefs (Shermer 2011). This explains the conviction of *active-anti* cohort towards their anti attitudes over long periods of time. A more concerning fact about conspiratorial beliefs is their "self-sealing quality" – attempts to reject the theory may backfire and those very attempts may be characterized as further proof of conspiracy (Sunstein and Vermeule 2009). Hence conspiratorial beliefs are very hard to correct. This suggests that dispelling vaccination myths among long-term anti-vaccine supporters might be very difficult, despite the amount of scientific evidence and rational arguments provided by government officials, scientific journals or other regulatory bodies (Wolfe 2002).

What about emotional appeals to dismiss vaccine myths? We find that complementing their cognitive concreteness with decreased *positive emotion* expressions, long term *anti* vaccine advocates can be identified as categorical thinkers – people whose writing is more focused on objects, things and categories, marked by higher use of nouns, articles and prepositions (Pennebaker 2013). Such categorical thinkers tend to be emotionally distant. In contrast, *active-pro* users characterized by higher cognitive processing and an informal expression style signal dynamic thinking (Pennebaker 2013). Thus, appealing to the emotional side of the anti-vaccination movement also is not likely to successfully change their attitudes.

Finally, long-term anti-vaccination advocates indicate a sense of group solidarity. Social psychologists have shown that higher *ingroup* language (like *member, family*) is indicative of group unanimity and a sense of shared identity (Tausczik and Pennebaker 2010). Long-term anti-vaccination attitude holders' increased usage of *ingroup* language, greater *attention-status* ratio and farther *message reach* are indicative of higher group cohesion, particularly in comparison to their pro-vaccination counterparts. Enhanced group cohesion and message reach are useful properties for establishing their beliefs as a movement and to recruiting new members into it.

In summary, long-term anti-vaccination supporters are resolute in their conspiracy worldviews, are categorical thinkers not likely to be appealed to by emotion, and exhibit high group cohesion. In light of these results, we need new tactics to counter the damaging consequences of anti-vaccination beliefs. Drawing from Sunstein's work on conspiracy theories, one possible strategy might be to introduce diversity in the closely-knit anti-vaccination groups (Sunstein and Vermeule 2009). Recall that anti-vaccination proponents display high *ingroup* characteristics. Weakening the well-knit cognitive clusters of extreme theories by introducing informational and social diversity (Sunstein and Vermeule 2009) might reduce the pool of long-term anti-vaccination advocates.

Predisposition among adoptees of anti-vaccine attitudes

Recall that new adoptees of anti-vaccine attitudes reveal themes indicating conspiracy thinking as well. Most importantly these themes are captured while tracking the user's entire twitter history and not just the posts after they start expressing anti-vaccine attitudes. This reveals a salient trait of *joining-anti* users: they see the world through the same paranoid lens as the *active-anti* users. In other words, they are already predisposed to adopt anti-vaccine sentiments. A predisposition is a state of an individual which when activated by a stimulus makes a person respond preferentially to the stimulus (Rokeach 1968).

The CDCwhistleblower controversy is an example of how a social media-driven event can trigger people's preexisting disposition to conspiracy thinking and subsequently provoke their attitudinal expressions that are counterproductive to society. Our current analysis cannot draw causal links from the event, but the timeline is highly suggestive: (a) people with prior indications of conspiratorial thinking start expressing anti-vaccine attitudes towards a topic after an event, where (b) the topic's anti-attitude is grounded in conspiracy theories and (c) the event itself is an alleged conspiracy event. Together this signals the power that a single event can exert on attitude formation. In this case the attitude is detrimental to society. Additional research to explore and generalize how events impact attitudes at population scale and how to control damaging attitude formation triggered by an event are fruitful areas for social media-based social science.

The linguistic analysis results show another characteristic of new adoptees of anti-vaccine attitudes: their lack of assuredness. Perhaps this lack of definitiveness nudges the predisposed minds of *joining-anti* cohort to latch on to the group of anti-vaccine advocates. We also find that they use more social expressions than their long-term counterparts, again a quality likely to make a person join a group (Gudykunst, Ting-Toomey, and Chua 1988).

Limitations & Future Work

Our user cohorts manifest the biases of a population of active social media users engaged in discussing vaccine and health information online. How representative their attitudes and opinions are of the general population is debatable (Keelan et al. 2010) and this confines the generalizability of our findings. Studies comparing users and nonusers of social media sites have found higher usage among young adults, no significant gender differences and lower percentage of Native American users (Hargittai 2007). Future work can examine cohorts who are not heavy social media users but still have strong anti-vaccination views. How do such individuals differ from those relying on social media to express vaccine opinions?

We were as thorough as possible in choosing the search phrases used to collect vaccine specific posts. However the transient nature of social media sites might surface new search terms not included in our analysis. Hence our results are not exhaustive. In this regard, we created a high precision rather than an all-inclusive but potentially noisy vaccine post dataset. Third, our analysis is based on quantitative observations rather than experimentation. Thus we can describe "what" we observe, but our "why" explanations are not causal. Future work, both qualitative and experimentation on why pro or anti vaccine advocates exhibit certain attitudes would help deepen the understanding of this phenomenon. Finally, our results exemplify vaccine attitudes on just one social media platform: Twitter. We do not know how this translates to other sites. We hope that future research can build up on our findings and investigate vaccine discussions or more general health information debates on other social media sites and mainstream media.

Conclusion

Through a case study of the vaccination debate, we demonstrate how analysis of the natural language expressions and social media activities can paint a multi-faceted picture of attitudes around a factious topic. Our study of principal actors in the vaccine debate in a key social media channel revealed that long-term anti-vaccination supporters are resolute in their beliefs and they tend toward categorical thinking and conspiratorial worldviews. New anti-vaccination adoptees share similar conspiracy thinking and hence are predisposed to develop attitudes aligned with the cohort of believers of vaccine myths. These new members tend to be less assured and more social in nature, but with a new and continued focus on health concerns. Our case study suggests that even a single event, CDCWhistleBlower in this case, may be a sufficient trigger for these people to adopt an attitude and join a socially counterproductive movement. Given these findings, new interventions such as trying to weaken the long-term government conspiracy base of the movement are needed.

References

Engelmore, R., and Morgan, A. eds. 1986. *Blackboard Systems*. Reading, Mass.: Addison-Wesley.

Abelson, Robert P. 1988. "Conviction." *American Psychologist* 43 (4): 267.

Allport, Gordon W., and Henry S. Odbert. 1936. "Trait-Names: A Psycho-Lexical Study." *Psychological Monographs* 47 (1): i.

Bem, Daryl J. 1970. "Beliefs, Attitudes, and Human Affairs.," 34.

"Bilderberg Group." 2015. Wikipedia, the Free Encyclopedia. https://en.wikipedia.org/wiki/Bilderberg_Group.

Boyd, R. L. 2015. "MEH: Meaning Extraction Helper (Version 1.3.02)[Software]." http://meh.ryanb.cc.

Boyd, Ryan L., Steven R. Wilson, James W. Pennebaker, Michal Kosinski, David J. Stillwell, and Rada Mihalcea. 2015. "Values in Words: Using Language to Evaluate and Understand Personal Values." In *Proc.* ICWSM.

Cantor, Nancy. 1990. "From Thought to Behavior: 'Having' and 'Doing' in the Study of Personality and Cognition." *American Psychologist* 45 (6): 735.

CDC. 2015. "Measles Cases in the United States Reach 20-Year High". http://cdc.gov/media/releases/2014/p0529-measles.html.

Chung, Cindy K., and James W. Pennebaker. 2008. "Revealing Dimensions of Thinking in Open-Ended Self-Descriptions: An Automated Meaning Extraction Method for Natural Language." *Journal of Research in Personality* 42 (1): 96–132.

Cohn, Michael A., Matthias R. Mehl, and James W. Pennebaker. 2004. "Linguistic Markers of Psychological Change Surrounding September 11, 2001." *Psychological Science* 15 (10): 687–93.

De Choudhury, Munmun. 2015. "Anorexia on Tumblr: A Characterization Study." In *Proc*. ICDH.

Dovidio, John F., and Russell H. Fazio. 1992. "New Technologies for the Direct and Indirect Assessment of Attitudes." Questions about Questions: Inquiries into the Cognitive Bases of Surveys 21: 204–37.

Eagly, Alice H., and Shelly Chaiken. 1993. The Psychology of Attitudes. Harcourt Brace Jovanovich College Publishers.

Eiser, J. Richard. 1975. "Attitudes and the Use of Evaluative Language: A Two-Way Process*." *Journal for the Theory of Social Behaviour* 5 (2): 235–48.

Gudykunst, William B., Stella Ting-Toomey, and Elizabeth Chua. 1988. Culture and Interpersonal Communication. Sage Pubs.

Hargittai, Eszter. 2007. "Whose Space? Differences among Users and Non-Users of Social Network Sites." *Journal of Computer-Mediated Communication* 13 (1): 276–97.

Hassan, Ahmed, Vahed Qazvinian, and Dragomir Radev. 2010. "What's with the Attitude?: Identifying Sentences with Attitude in Online Discussions." In *Proc*. EMNLP.

Holman, E. Alison, and Roxane Cohen Silver. 1998. "Getting' Stuck' in the Past: Temporal Orientation and Coping with Trauma." *Journal of Personality and Social Psychology* 74 (5): 1146.

Hutto, C. J., Sarita Yardi, and Eric Gilbert. 2013. "A Longitudinal Study of Follow Predictors on Twitter." In *Proc*. CHI.

Kata, Anna. 2010. "A Postmodern Pandora's Box: Anti-Vaccination Misinformation on the Internet." *Vaccine* 28 (7): 1709–16.

Keelan, Jennifer, Vera Pavri, Ravin Balakrishnan, and Kumanan Wilson. 2010. "An Analysis of the Human Papilloma Virus Vaccine Debate on MySpace Blogs." *Vaccine* 28 (6): 1535–40.

Kramer, Adam DI, and Cindy K. Chung. 2011. "Dimensions of Self-Expression in Facebook Status Updates." In *Proc.* ICWSM.

Kristiansen, Connie M., and Mark P. Zanna. 1988. "Justifying Attitudes by Appealing to Values: A Functional Perspective." *British Journal of Social Psychology* 27 (3): 247–56.

Markus, Hazel. 1977. "Self-Schemata and Processing Information about the Self." *Journal of Personality and Social Psychology* 35 (2): 63.

Newman, Matthew L., James W. Pennebaker, Diane S. Berry, and Jane M. Richards. 2003. "Lying Words: Predicting Deception from Linguistic Styles." *Personality and Social Psychology Bulletin* 29 (5): 665–75.

O'Connor, Brendan, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. "From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series." In. ICWSM.

Pennebaker, James W. 2013. The Secret Life of Pronouns: What Our Words Say About Us. Reprint edition. Bloomsbury Press.

Petty, Richard E., and Jon A. Krosnick. 2014. Attitude Strength: Antecedents and Consequences. *Psychology Press*.

Pomerantz, Eva M., Shelly Chaiken, and Rosalind S. Tordesillas. 1995. "Attitude Strength and Resistance Processes." *Journal of Personality and Social Psychology* 69 (3): 408.

Rokeach, Milton. 1968. "Beliefs, Attitudes and Values: A Theory of Organization and Change."

Rude, Stephanie, Eva-Maria Gortner, and James Pennebaker. 2004. "Language Use of Depressed and Depression-Vulnerable College Students." *Cognition & Emotion* 18 (8): 1121–33.

Salmon, Daniel A., Lawrence H. Moulton, Saad B. Omer, M. Patricia deHart, Shannon Stokley, and Neal A. Halsey. 2005. "Factors Associated with Refusal of Childhood Vaccines among Parents of School-Aged Children: A Case-Control Study." *Archives of Pediatrics & Adolescent Medicine* 159 (5): 470–76.

Savion, Leah. 2012. "Clinging to Discredited Beliefs: The Larger Cognitive Story." *Journal of the Scholarship of Teaching and Learning* 9 (1): 81–92.

Schuman, Howard, and Stanley Presser. 1981. Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context. Sage.

Schwarz, Norbert. 2007. "Attitude Construction: Evaluation in Context." *Social Cognition* 25 (5): 638–56.

Shermer, Michael. 2011. The Believing Brain: From Ghosts and Gods to Politics and conspiracies—How We Construct Beliefs and Reinforce Them as Truths. Macmillan.

Slatcher, Richard B., Cindy K. Chung, James W. Pennebaker, and Lori D. Stone. 2007. "Winning Words: Individual Differences in Linguistic Style among US Presidential and Vice Presidential Candidates." *Journal of Research in Personality* 41 (1): 63–75.

Slobin, Dan I. 1996. "From thought and language to thinking for speaking." In Rethinking Linguistic Relativity, 70–96. 17. Cambridge University Press.

Sunstein, Cass R., and Adrian Vermeule. 2009. "Conspiracy Theories: Causes and Cures*." *Journal of Political Philosophy* 17 (2): 202–27.

Swami, Viren, Rebecca Coles, Stefan Stieger, Jakob Pietschnig, Adrian Furnham, Sherry Rehim, and Martin Voracek. 2011. "Conspiracist Ideation in Britain and Austria: Evidence of a Monological Belief System and Associations between Individual Psychological Differences and Real-World and Fictitious Conspiracy Theories." *British Journal of Psychology* 102 (3): 443–63.

Tausczik, Yla R., and James W. Pennebaker. 2010. "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods." *Journal of Language and Social Psychology* 29 (1): 24–54.

Wolfe, Robert M. 2002. "Vaccine Safety Activists on the Internet." *Expert Review of Vaccines* 1 (3): 249.

Zonis, Marvin, and Craig M. Joseph. 1994. "Conspiracy Thinking in the Middle East." *Political Psychology*, 443–5.