Perceptions of Presidential Candidates' Personalities in Twitter

Sanmitra Bhattacharya, Chao Yang, and Padmini Srinivasan
Department of Computer Science, The University of Iowa, 14 MacLean Hall, Iowa City, IA 52242. E-mail: {sanmitra-bhattacharya, chao-yang, padmini-srinivasan}@uiowa.edu

Bob Boynton
Department of Political Science, The University of Iowa, 341 Schaeffer Hall, Iowa City, IA 52242. E-mail: bob-boynton@uiowa.edu

Political sentiment analysis using social media, especially Twitter, has attracted wide interest in recent years. In such research, opinions about politicians are typically divided into positive, negative, or neutral. In our research, the goal is to mine political opinion from social media at a higher resolution by assessing statements of opinion related to the personality traits of politicians; this is an angle that has not yet been considered in social media research. A second goal is to contribute a novel retrieval-based approach for tracking public perception of personality using Gough and Heilbrun's Adjective Check List (ACL) of 110 terms describing key traits. This is in contrast to the typical lexical and machine-learning approaches used in sentiment analysis. High-precision search templates developed from the ACL were run on an 18-month span of Twitter posts mentioning Obama and Romney and these retrieved more than half a million tweets. For example, the results indicated that Romney was perceived as more of an achiever and Obama was perceived as somewhat more friendly. The traits were also aggregated into 14 broad personality dimensions. For example, Obama rated far higher than Romney on the Moderation dimension and lower on the Machiavellianism dimension. The temporal variability of such perceptions was explored.

Introduction

As more and more people have adopted social media, public communication about politics has exploded. Twitter is particularly important in this change because of its scale: About a billion tweets are posted every 5 days and there are close to 650 million users.[1] This communication has an impact on how people think about candidates and thus on the outcome of an election. Although perceptions of the personalities of candidates have been considered important for much of the history of electoral research, the issue has been difficult to research systematically. On Twitter, however, individuals, in large numbers, express their perceptions of the candidates. Changes can be tracked over time. If the messages are analyzed carefully, it may be possible to obtain some understanding of how the candidates are viewed.

Research Objectives

Our research makes several contributions to the study of social media and political analysis. First, Twitter messages are examined to assess perceptions of the personalities of candidates. Other studies sort messages about candidates into positive, neutral, and negative sentiment (Chung & Mustafaraj, 2011; Diakopoulos & Shamma, 2010; Gayo-Avello, Metaxas, & Mustafaraj, 2011), but in this research a well-established trait-and-personality lexicon of 110 terms is used that are then shown to form 14 dimensions. This gives both a fuller and a more fine-grained analysis of how political personalities are viewed. To the best of our knowledge this is the first study of social media to examine views of the personalities of candidates.

Second, the Twitter messages on personality perceptions are analyzed over time. This is to determine if the trends observed are consistent with intuitive expectations, given external events. For example, we examine trends around Obama's announcement of the auto bailouts in February 2012 and the October 2012 presidential debate.

Third, the method for extracting opinions on personality differs from the norms in political sentiment research, which we believe is an important contribution. A classifier-based approach (e.g., Monti et al., 2013; Wang, Can, Kazemzadeh,

[1] http://www.statisticbrain.com/twitter-statistics/

Received December 27, 2013; revised June 21, 2014; accepted June 23, 2014

 • Published online 3 April 2015 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.23377

Bar, & Narayanan, 2012; Younus et al., 2011) would be highly challenging, as it would require a classifier for each of 110 traits. A standard lexicon-based count method (e.g., Chung & Mustafaraj, 2011; Gayo-Avello et al., 2011; Kim & Yoo, 2012; Tumasjan, Sprenger, Sandner, & Welpe, 2010; Wu, Wong, Deng, & Chang, 2011) is risky in terms of accuracy. Instead, a high-precision retrieval strategy built from a trait-and-personality lexicon widely used in political science research is developed. This high-precision retrieval strategy is assessed for precision, recall, and *F* score (described later).

The research is based on an 18-month collection of Twitter messages mentioning Obama and Romney during the 2012 election campaign. Among the highlights of our findings: Obama is perceived by Twitter users as moderate, intellectually brilliant, and a pacifist. Romney is perceived as having high achievement drive and being Machiavellian as well as inflexible. Public perceptions mark both individuals as conservative, witty, and physically attractive. These characterizations take public perceptions of the candidates well beyond positive, neutral, and negative. With temporal analysis we find, for example, a big spike around the October presidential debate for Obama in the Moderation dimension when he emphasized moderation in various internal and external affairs.

In summary, this study is a mining of public perceptions of the personalities of the two U.S. presidential candidates in the 2012 elections using Twitter. The study contributes both to the stream of social media research on mining political sentiment and to electoral research in political science. In social media research, the study is distinct in its fine-grained focus, its use of a new lexicon, and the use of high-precision retrieval. In electoral research, it addresses an area that has been difficult to study systematically.

In the next section we present a selective literature review covering areas such as the importance of personality perceptions in political science research, social media research on elections and politicians, sentiment analysis research, and the use of lexicons. Then, we present our methods introducing the ACL trait lexicon, personality dimensions, tweet retrieval strategy, and validation. This is followed by presentation of our results gathered from our observations about traits and personality dimensions and from our temporal analysis. This section also includes comparisons with Gallup and Pew observations. The last section presents our conclusions, including plans for future work.

Literature Review

From the 1960s onward, political science research on public opinion and elections has made it clear that the primary factors in choice of a candidate are (a) partisanship, (b) the issues of the campaign, and (c) impressions of the person running for office (Campbell, Converse, Miller, & Stokes, 1960). Of the three, impressions of the candidates have been the most difficult to study. By examining postings of personality-related opinion on Twitter a major advance in this direction is possible. By being able to systematically compare perceptions of candidates and by noting changes over time the weakest of the three factors can now become a strength in research.

Personality traits of politicians have been assessed to study their influence on leadership and decision-making styles (Costantini & Craik, 1980; Gallagher & Allen, 2014; George & George, 1998; Kinder, 1978). For example, Gallagher and Allen (2014, p. 1) find that leaders with an "excitement seeking" personality tend to use force to meet their goals whereas those who are "open to action" tend to show greater variance in their decision making. Overall, they find that the personality traits of presidents shape their choices and their level of consistency in policy making. However, this prior body of research examines the personalities of political leaders rather than the *perceptions* of the personalities of leaders by the public.

Public perception of the personality of political leaders may vary widely across a population and even over time. The Pew Research Center for the People & the Press in conjunction with *The Washington Post* has made a start on researching this variation (Pew Research Center for the People & the Press, 2012a, 2012b). The Center interviewed U.S. adult voters multiple times over the phone, asking for the one word that comes to mind when a politician's name is mentioned. Some of the words generated concern personality traits. In September 2008, the most frequent one-word response for Obama was the trait *inexperienced* whereas in August 2012 the response was *being good* (as a man and at the job). However, this is a very limited examination of perceptions of personality.

A number of social media studies have centered on politicians, elections, and other political events. The table in the Appendix provides a summary of 41 papers reviewed. The review is not comprehensive, but focuses on a core of social media papers about politics with an emphasis on sentiment analysis. Around 50% of the papers estimate the level of support or approval for one or more politicians (e.g., Mejova, Srinivasan, & Boynton, 2013; Mohammad, Kiritchenko, & Martin, 2013), largely using sentiment analysis (e.g., Bravo-Marquez, Gayo-Avello, Mendoza, & Poblete, 2012; Diakopoulos & Shamma, 2010; Gayo-Avello et al., 2011). Sometimes this is done using counts of occurrences of keywords and hashtags related to the politician's name (e.g., Contractor & Faruquie, 2013; Hanna et al., 2013; Jungherr, 2013). The count approach is a simple one; the idea is to get a normalized count of the number of tweets containing a relevant keyword or a relevant hashtag. Machine-learning classifiers have also been used to gauge political support. For example, Mohammad et al. (2013) used classifiers to determine the intent behind a political tweet toward Obama and Romney in the 2012 U.S. presidential elections. They explored 11 categories for purposes such as agree, praise, support, and disagree.

There has been significant enthusiasm for using estimated level of support from Twitter to predict election outcomes and approval ratings, with some successes at

DOI: 10.1002/asi

predicting outcomes (e.g., Gaurav, Srivastava, Kumar, & Miller, 2013; Sang & Bos, 2012; Soler, Cuartero, & Roblizo, 2012) and some failures (e.g., Chung & Mustafaraj, 2011; Jungherr, 2013; Mejova & Srinivasan, 2012). Election outcome prediction is also an area that has generated a few critiques (e.g., Gayo-Avello, 2012; Mejova et al., 2013). Of particular note are the observations regarding differences between the "silent majority" and the "vocal minority" on Twitter (Mustafaraj, Finn, Whitlock, & Metaxas, 2011, p. 103).

Several papers focused on understanding the nature of Twitter political communications. Al-Khalifa (2011) used hashtags to visualize the shape of political networks. Starbird and Palen (2012) studied information diffusion in the context of the 2011 Egyptian protests. Eveland and Hutchens (2013) used social network data from student groups to study the role of conversations in forming political perceptions. Stieglitz and Dang-Xuan (2012) showed that sentiment affects retweet rates in the political domain. Kim and Yoo (2012) demonstrated that the degree of emotion expressed affects the number of replies and retweets.

The papers most directly related to our work are those on sentiment analysis of political tweets. Most consider three sentiment classes of positive, negative, and neutral (e.g., Jungherr, 2013; Nooralahzadeh, Arunachalam, & Chiru, 2013; Wang et al., 2012). A few look at other sentiment aspects such as anxiety, anger, and sadness (Tumasjan et al., 2010) or categories such as agree, praise, support, and disagree (Mohammad et al., 2013). As already noted, none of them examines personality traits. The majority use lexicon-based methods, such as LIWC (e.g., Kim & Yoo, 2012; Tumasjan et al., 2010), OpinionFinder Lexicon (e.g., Chung & Mustafaraj, 2011; Gayo-Avello et al., 2011), and SentiStrength (e.g., Wu et al., 2011).

This is also the case for the general sentiment literature (e.g., Thelwall, Buckley, & Paltoglou, 2012). Several papers use supervised classifiers (e.g., Monti et al., 2013; Wang et al., 2012; Younus et al., 2011). This trend reflects what is seen in the general sentiment analysis literature (e.g., Pang & Lee, 2008; Pang, Lee, & Vaithyanathan, 2002; Pak & Paroubek, 2010). Hybrid approaches that combine the two for sentiment analysis have also been used (e.g., Jansen, Zhang, Sobel, & Chowdury, 2009; Melville, Gryc, & Lawrence, 2009). For example, in the general literature, Jansen et al. (2009) use a multinomial Bayes classifier and a lexicon of words or phrases each with a probability of being either a positive or negative word.

There is no prior research using high-precision retrieval strategies as the basis for finding relevant tweets conveying political sentiment. Our research uses a widely accepted list of personality traits known as the Adjective Check List (ACL) developed by Gough and Heilbrun (Gough & Heilbrun, 1983). The approach may also appear to be lexical in nature, albeit with a new lexicon, but there is a crucial difference. Specifically, the ACL is used to build a set of high-precision query templates to retrieve relevant Twitter posts. The ACL has not only been highly cited since its inception, it has also been widely accepted as a commercial tool for assessment of psychological traits. Although the ACL has been used extensively in a broad range of psychological studies ranging from presidential personalities (Simonton, 1986) to employee–organization fit (O'Reilly, Chatman, & Caldwell, 1991), it has not been used to analyze expressions of public perception of personality traits of presidential candidates from Twitter or any other social media.

The ACL traits are not the only classification scheme relating to personality as the assessment of human personality has a long history in psychology. The "lexical hypothesis" that most significant personality characteristics are encoded in people's natural language has been around for more than a century (Caprara & Cervone, 2000) and has motivated the development of various trait lists, categories, and dimensions. Among the earliest we have the work of Allport and Odbert (1936) who extracted close to 18,000 personality-related terms from an unabridged dictionary. Refinements and categorizations over the years eventually led to the "Big Five" personality dimensions (or factors): Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness (Tupes & Christal, 1992). Each dimension represents personality at the broadest level, summarizing a large number of distinct, more specific personality characteristics. For this research on mining public perceptions of personality traits, the Big Five classes are more abstract than desired.

The LIWC software (and categories) was developed to study the emotional, cognitive, and structural components present in individuals' verbal and written speech samples (Tausczik & Pennebaker, 2010). Using a lexicon to map individual words into categories, it outputs counts on about 80 variables that cover 26 linguistic (pronouns, verbs and their tenses, negation, etc.), 32 psychological (family, friends, positive emotion negative emotion, certainty, etc.), seven personal (work, achievement, money, etc.), and three spoken (assent, nonfluencies, fillers) aspects. Studies have shown that some of the LIWC categories correspond with the Big Five personality traits. For example, higher word counts, fewer large words, and lower language complexity predicted Extraversion (Tausczik & Pennebaker, 2010). In personality studies LIWC has been used primarily to gauge the personality of a writer (through writing samples) or a speaker (through dialogue samples). LIWC is not designed to gauge perception of personality of third persons. Also, compared to the ACL, which is a direct list of trait adjectives, one would have to infer traits from LIWC outputs.

Methods

Gough and Heilbrun's Adjectives of Personality Traits

The ACL developed by Gough and Heilbrun (1983) consists of 300 adjectives and phrases used to describe an individual's personality. Examples include *tough*, *cheerful*, *coarse*, and *spontaneous*. The ACL has been used

DOI: 10.1002/asi

TABLE 1. Fourteen personality dimensions and sample traits with loadings.

Personality dimension	No. of positive traits	No. of negative traits
Moderation	21 (e.g., considerate, 0.63)	37 (e.g., blustery, –0.85)
Friendliness	21 (e.g., easy going, 0.73)	17 (e.g., suspicious, –0.51)
Intellectual Brilliance	10 (e.g., intelligent, 0.64)	2 (e.g., dull, –0.71)
Machiavellianism	6 (e.g., deceitful, 0.87)	2 (e.g., honest, –0.63)
Poise and Polish	6 (e.g., sophisticated, 0.62)	5 (e.g., coarse, –0.46)
Achievement Drive	2 (e.g., persistent, 0.76)	3 (e.g., effeminate, –0.69)
Forcefulness	5 (e.g., energetic, 0.64)	0
Wit	4 (e.g., humorous, 0.74)	0
Physical Attractiveness	3 (e.g., attractive, 0.62)	0
Pettiness	2 (e.g., greedy, 0.68)	0
Tidiness	3 (e.g., organized, 0.7)	1 (courageous, –0.4)
Conservatism	2 (e.g., conventional, 0.55)	0
Inflexibility	4 (e.g., persistent, 0.43)	0
Pacifism	1 (peaceable, 0.61)	1 (courageous, –0.48)

in many studies. In a 1986 paper that is more directly related to this research, Simonton adapted the ACL to manually compare the personalities of 39 American presidents using biographical sentences. For this he reduced the 300 adjectives to a set of 110 that could be reliably measured. He found, for example, that T. Roosevelt and Jackson were the least moderate of the 39 presidents, whereas Jefferson was perceived as the most intellectually brilliant.

Trait Aggregation Into Personality Dimensions

Using factor analysis, Simonton (1986) also grouped the 110 adjectives into 14 nontrivial personality dimensions. Each component trait has a numerically rated "loading" between −1 and 1. A positive loading indicates that the trait directly contributes to the personality dimension—that is, has a direct relationship. If negative, then the presence of the trait reduces the strength of that personality dimension. The magnitude indicates the (relative) extent of contribution/detraction. The 14 dimensions in order of the number of adjectives contained—that is, dominance of factors—are shown in Table 1.

Note that a trait may be in more than one dimension. For example, *courageous* is in Pacificism and in Tidiness. Also, some of these categorizations may be a bit doubtful given current usage. Slight adjustments are made and described at the appropriate points in this article.

For use with the Twitter messages the list is augmented by identifying synonymous and antonymous words. Nine hundred and two synonyms and 989 antonyms were added (average of 8.2 and 9 per trait, respectively). These were identified using WordNet 3.07 and then manually filtered to remove terms that may not be applicable to individuals. The full list of trait synonyms and antonyms used is available upon request from the authors.

Positive, Negative, and Neutral Traits

With close to three decades since the publication of the ACL, interpretations of some of the traits have evolved. For example, *shrewd* and *courageous* were rated as negative, and *meek* and *evasive* were rated as positive. Current interpretations were obtained through an exercise with 11 Amazon Mechanical Turk[2] workers, who were hired to provide ratings of positive, negative, or neutral for each of the 110 traits. They were asked: "When evaluating a candidate for the Presidency of a country (or for a similar country-level Leadership position) do you regard this trait as representing a positive, negative or neutral characteristic?" Each worker (with at least a high-school degree) was paid a fixed amount of $2.00 for labeling all 110 traits. A majority vote[3] was used; interannotator agreement achieved was 66%, which is lower than preferred (70%). Examining the ratings, we find that for around 25% of the traits one class received all the votes (e.g., *intelligent* and *assertive* received all positive votes while *confused* and *deceitful* received all negative votes). For 94% of the traits, one class received more than 50% of the total votes (e.g., *cautious*: 3 neutral votes, 7 positive, and 1 negative). *Aggressive* (5 positive, 5 negative, and 1 neutral) and *humorous* (5 positive, 1 negative, and 5 neutral) were the only two traits of the 110 without a clear majority vote. We ignore these two traits in our analysis.

Each Trait a Continuum

We view each trait as being associated with a continuum of values. For example, for the trait *friendly* one can visualize a continuum with "is friendly" at one end and its antonym "is unfriendly" at the opposite end (see Figure 1). One can then logically map tweets to different points on the continuum. Clearly the tweet "Mary is really an unfriendly person" is close to, if not at, the "is unfriendly" end. "Mary is a somewhat friendly person" is closer to the "is friendly" end. Tweets of interest include those indicating the presence or absence of a trait and also those indicating the presence or absence of the trait's antonym. Table 2 illustrates this point. Each cell holds a combination of interest. When gauging the perceived friendliness of a person we would like to find tweets that fall into cells (1,1) and (2,2) whereas for unfriendliness we want tweets that fall into cells (1,2) and (2,1).

Additionally note that Statement 1 is stronger than Statement 2, which, in turn, is stronger than Statement 3. Similar gradations are seen in Statements 4, 5, 6, and so on. Therefore, tweets that indicate Mary is really friendly would be posts like Statement 1. However, tweets simply saying Mary is friendly would be statements like 1 and 2 or 1 through 3. Statements claiming the absence of the antonym (Statements

[2]https://www.mturk.com/mturk/

[3]In case we have a tie for majority vote (e.g., 5 positive, 5 negative, and 1 neutral rating), we do not mark the trait as positive, negative, or neutral. There were two such cases: "aggressive" and "humorous."

DOI: 10.1002/asi

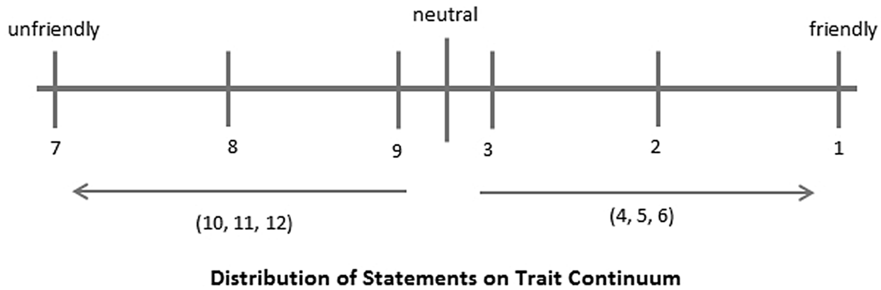

FIG. 1. Sample trait continuum.

TABLE 2. Statement types.

	Trait (Friendly)	Trait antonym (Unfriendly)
Present	1. Mary is very friendly.	7. Mary is really unfriendly.
	2. Mary is friendly.	8. Mary is unfriendly.
	3. Mary is somewhat friendly.	9. Mary is somewhat unfriendly.
Absent	10. Mary absolutely isn't friendly.	4. Mary absolutely isn't unfriendly.
	11. Mary isn't friendly.	5. Mary isn't unfriendly.
	12. Mary is kind of not friendly.	6. Mary kind of isn't unfriendly.

TABLE 3. Data set characteristics.

	Obama data set	Romney data set
No. of tweets	81,200,065	41,860,086
No. of retweets	35,390,749 (44%)	24,234,388 (58%)
No. with hashtags	23,399,894 (29%)	12,403,064 (30%)
No. with URL	40,354,966 (50%)	20,452,977 (49%)
No. of users	12,629,969	5,650,921

4, 5, and 6) also indicate some degree of friendliness. They may not specify the degree of friendliness, but such tweets would add to the count of "is friendly" tweets rather than to the "is very friendly" set.

Figure 1 summarizes these observations by plotting the 12 types of statements about the trait *friendly* in Table 2 on a continuum. What is important is the relative positioning of points; the distance between them is not the focus. Note that Statements 1, 2, 3, 7, 8, and 9 may be readily plotted with respect to each other. In contrast Statements 4, 5, and 6, from cell (2,1), suggest an opinion somewhere on the side of friendliness and therefore fall in a range. Similarly, Statements 10, 11, and 12 fall in a range on the unfriendliness side.

Measures

Equations 1 and 2 define measures for estimating the presence of a trait for an individual.

$$TraitPresenceScore = \alpha \times [1] + \beta \times [2] + \gamma \times [3] + \delta \times ([4] + [5] + [6]) \quad (1)$$

$$TraitAbsenceScore = \lambda \times [7] + \mu \times [8] + \nu \times [9] + \xi \times ([10] + [11] + [12]) \quad (2)$$

In the above, for example, "[5]" refers to the number of tweets containing information corresponding to Statement 5 of Table 2. The interpretation is analogous for [1], [2], . . . ,[12]. A weighted sum is used to rate strong statements higher. For simplicity, the weights 4, 3, 2, and 1 are used for α, β, γ, and δ. And −4, −3, −2, and −1 are used as weights for λ, μ, ν, and ξ, respectively. Using this strategy, identical distributions of tweets indicating presence and absence of a trait give the same scores on both sides but with opposite signs. This would indicate that the tweeters are evenly split in their opinion over that trait and the summary reading is one of neutrality. Also if the distribution of tweets is proportionate on both sides, but the raw number of tweets is higher on one side then that side will have a higher score (ignoring sign).

Comparing perceptions of individuals whose tweet collections vary in size requires normalizing the scores. In the data set there are twice as many tweets about Obama as about Romney (see Table 3), which requires some form of normalization. Assume that there are a total *M* tweets for a trait on an individual. Then the two extreme scores occur if all of them fall in set [1] or all fall in set [7] of Table 2. The score is then $4 \times M$ or $-4 \times M$, respectively. The neutral value is 0 when *TraitPresenceScore* and *TraitAbsenceScore* are equal. The normalized score for the trait is computed as

$$TraitScore = \frac{TraitPresenceScore + TraitAbsenceScore}{4 \times M} \quad (3)$$

This score represents the distance from the neutral on a scale of −1 to +1. If *TraitScore* is +1, for example, all tweets are in set [1]. This measure is agnostic to the number of tweets found for a trait. When a tweet set retrieved is small the computation may give readings that would not be stable.

The estimate of perceptions of each of the 14 personality dimensions is a weighted sum over the component traits. The weights are the loadings assigned to the traits (see the

DOI: 10.1002/asi

TABLE 4. Examples of search templates, *P* standards for the person (Obama or Romney).

Search template	Example statement retrieved	Examples of word classes
1. [*P*] is [A] [*T*]	1. Obama is certainly smart.	[A]: truly, certainly, definitely (similar adverbs) *T*: trait
2. [*P*] is ?[Q] [*T*]	2. Romney is quite smart.	[Q] (optional): quite, also, even, sufficiently, . . .
3. [*P*] is [S] [*T*]	3. Obama is sort of smart.	[S]: somewhat, sort of, kinda, almost, . . .
4. [*P*] is [A] not [ant(*T*)]	4. Romney is definitely not dumb.	[I]: person, individual, leader . . .
5. [*P*] is not [antonym (*T*)]	5. Obama is not dumb.	
6. [*P*] is [S] not [antonym (*T*)]	6. Romney is sort of not dumb.	
7. [*P*] is a/an [A] [*T*] [I]	7. Romney is a truly smart person.	
8. [*P*] is ?[Q] a/an [*T*] [I]	8. Obama is also truly a smart individual.	
9. [*P*] is [S] a/an [*T*] [I]	9. Obama is somewhat a smart person.	
10. [*P*] is [A] not a/an [antonym(*T*)] [I]	10. Romney is definitely not a dumb person.	
11. [*P*] is not a/an [antonym (*T*)] [I]	11. Obama is not a dumb individual.	
12. [*P*] is [S] not a/an [antonym (*T*)] [I]	12. Romney is somewhat not a dumb person.	

Note. *T* = one of the 110 traits studied. "?" indicates optional. Column 2 uses "smart" as the example trait with "dumb" as an example antonym.

prior section on Trait Aggregation). Thus for a given personality dimension made up of *n* traits,

$$DimensionScore = \sum_{i=1}^{n} (Loading_i \times TraitScore_i) \qquad (4)$$

Because each dimension has a different number of component traits and the loadings on each trait appear somewhat independent, the range of *DimensionScore* is variable across the 14 dimensions. To make *DimensionScores* comparable across dimensions the scores are normalized. Because *TraitScores* fall in the range of +1 to −1, the maximum *DimensionScore* occurs when *TraitScores* are +1 for all component traits with positive loadings and are −1 for all traits with negative loadings. The opposite is true for the minimum *DimensionScore*. Thus the two extremes are in the range of [*MaxDimensionScore*, *MinDimensionScore*] where *MaxDimensionScore* = (Σ*Positive Loadings* − Σ*Negative Loadings*) and *MinDimensionScore* = (−Σ*Positive Loadings* + Σ*Negative Loadings*). These two extremes are identical in magnitude and differ only in sign. For example, for the Moderation personality dimension the range is [35.5, −35.5] and for Machiavellianism it is [5.13, −5.13]. The neutral value in each case is 0 when the scores on the positive and negative sides are even. A score of 0 indicates that, when taken in aggregate, Twitter users do not perceive the personality dimension as present in the individual nor do they perceive the dimension as absent in the individual. We make use of the neutral score to normalize the *DimensionScore* as follows:

$$NormDimensionScore = \frac{DimensionScore}{MaxDimensionScore} \qquad (5)$$

This score represents the distance of the *DimensionScore* from the neutral on a scale of −1 to +1. A score of +1 (−1) indicates that, as per the collection of relevant tweets, the individual is given the highest (lowest) rating for that personality dimension. This score is also agnostic to the size of the relevant tweet set, and again the reading may not be stable when the set is small.

Identifying Personality Statements in the Twitter Messages

The previous section laid out procedures for measuring public perception of personality traits and dimensions. However, it did not indicate how to systematically identify relevant statements of perception of personality from the Twitter collection. That is a critical step in our research, and a step for which we offer different procedures than have been widely used in prior sentiment analysis research.

Finding the tweets discussing each personality trait for either politician using a classifier-based approach would be highly challenging, as a classifier would be needed for each trait. That would involve a large amount of training data overall. A standard lexicon-based count method is an approximation and so risks high false positives. For example, the tweet "Obama has a strong, a really strong vice prez" has a high count for the trait *strong* but it is clearly not being used with reference to Obama.

Instead, a high-precision, template-based retrieval approach is used. The templates were built manually using the ACL. The goal during the process of building the search templates was to retrieve tweets conveying personality perceptions with high precision rather than to be comprehensive about retrieving relevant tweets. The statement types in Table 2 were used as a guideline during development. There are 40 templates; several examples are shown in Table 4. Each template specifies a general search strategy using variables. For example, the first template ([*P*] is [A] [*T*]) retrieves tweets with the phrase "Obama is certainly smart" or the phrase "Romney is absolutely intelligent." Here [*P*] is a variable for the person name, [A] is a word class, and variable [*T*] represents a trait. The third column gives examples of synonymous terms that are fillers for the template variables. These template-driven, high-precision searches were executed against indexes built using the Obama and Romney tweet collections. We use Indri (Strohman, Metzler, Turtle, & Croft, 2005) for indexing and retrieval.

Specifically, the 40 templates were applied to each trait for each politician, yielding a total of 8800 phrase queries (40 templates * 110 traits * 2 politicians). The search templates were designed to capture the key, direct expressions of opinion on a personality trait. It is possible that the templates miss other expressions as we have deliberately emphasized high precision. However, these omissions do not a priori express a bias toward either individual being compared. The template collection also includes variants for reductions such as "isn't" (not shown in the table). Additionally, the templates map directly onto the statement categories shown in Table 2. For example, template 7 belongs in cell (1,1) and corresponds to Statement 1. The full set of 40 trait-independent templates is available from the authors.

Tweet Collection

The Twitter data were collected using the Twitter Search API, with a search being conducted every 5 minutes for keywords such as "obama," "barackobama," "romney," and "mittromney." This was done from June 16, 2011 through November 16, 2012 (unforeseen incidents caused failure in retrieval on August 30 and 31, 2012). The tweets were filtered to remove those mentioning other individuals such as Michelle Obama and Ann Romney. Table 3 describes the overall characteristics of the two data sets. Although there are about double the number of Obama tweets as Romney tweets, the two collections seem fairly comparable. Romney's set has a higher tendency to include retweets. The number of tweets per user for Romney (7.4) is also a little higher compared to Obama (6.4).

Validation of Methods

Exogenous validation of the methods would be ideal. But there are obviously no gold standards with which to compare the results. Moreover, it is unrealistic to expect to find data sets of fixed personality ratings for any individual, as perceptions can vary over time. Later the findings in this research are compared with Pew surveys, support trends from the Gallup poll, and to the outcome of the election. But these comparisons do not deal specifically with the challenge of locating perceptions of personality in Twitter messages. That is the challenge leading to the decision to conduct *endogenous* validation of the methods.

For the *endogenous* evaluation the question is this: Are these the right sets of tweets for the observations of personality? A small-scale experiment was conducted to evaluate the merit of the template-based, precision-driven methods by first building a gold standard annotated data set.

To create the gold standard annotated data set, we first identified the five traits that retrieved the most tweets when combining retrieved sets for Obama and Romney. These traits are *confused*, *courageous*, *dull*, *easy going*, and *greedy*. We developed an unconstrained strategy as a baseline retrieval strategy. Specifically, we retrieved all tweets that mentioned either individual and one of the selected trait words. This baseline method is similar in spirit to the lexicon-based count methods in sentiment analysis research (e.g., Tumasjan et al., 2010; Kim & Yoo, 2012). Since this strategy relies solely on the presence of particular words, precision may be at risk. For example, the tweet "Isn't it paradoxical that Obama is not doing so badly in polls despite a weak jobs market report" contains the trait *weak* along with "Obama" but it is not associated with Obama's personality.

TABLE 5. Precision (**P**), **R**ecall (*R*), *F* score (**F**) for precision-driven and unconstrained baseline retrieval methods.

	Precision-driven retrieval	Unconstrained retrieval
Obama	**P** = 0.95, **R** = 0.35, **F** = 0.51	**P** = 0.13, **R** = 1.00, **F** = 0.23
Romney	**P** = 0.91, **R** = 0.46, **F** = 0.61	**P** = 0.19, **R** = 1.00, **F** = 0.32

From the tweets retrieved by the baseline strategy we randomly drew a sample of 200 tweets indicating trait presence and another 200 indicating trait absence for each candidate (using the trait, its synonyms, its antonyms, and negation words). This yielded a total of 800 sampled tweets for manual annotation. Each tweet was labeled for relevance with respect to a particular personality trait *and* its association with a politician.

The crowd-sourcing platform oDesk[4] was used for this purpose. Out of a pool of 12 candidates, three were selected for the annotation task based on their performance on an initial competency test. On average, annotators were paid 1.5 cents/tweet and it took them around one week to complete the task. Majority vote was taken for the judgment of each tweet. The kappa score for interannotator agreement was 69%. This is just short of the 70% score that is generally accepted as reasonable.

This annotated data set of 800 tweets was used to compute recall, precision, and *F* score (harmonic mean of recall and precision) of the template-driven strategy and for the unconstrained baseline strategy. Table 5 presents the results for the Romney and Obama subsets. Although recall is at 1.0 for the unconstrained baseline method, precision is extremely poor (highest score is 0.19). The template search strategy favors precision over recall as expected. This preference is intentional given the extreme level of noise in Twitter data. *F* scores for the precision-driven methods are clearly superior and quite reasonable given that retrieval from social media is a hard problem. Improvements in *F* scores are 121% and 91% compared to the baseline for the Obama and Romney data, respectively. This shows the suitability of the template method for identifying and measuring personality traits versus other unconstrained retrieval methods. Future research will focus on improving recall using this gold standard data. Key to note at this point is that the search strategies were the same for each individual and so there are no identifiable biases in our methods.

Error analysis indicates that, as previously mentioned, most of the false positive errors occurred due to two reasons. First, negations in tweets asking questions were not

[4]https://www.odesk.com/

DOI: 10.1002/asi

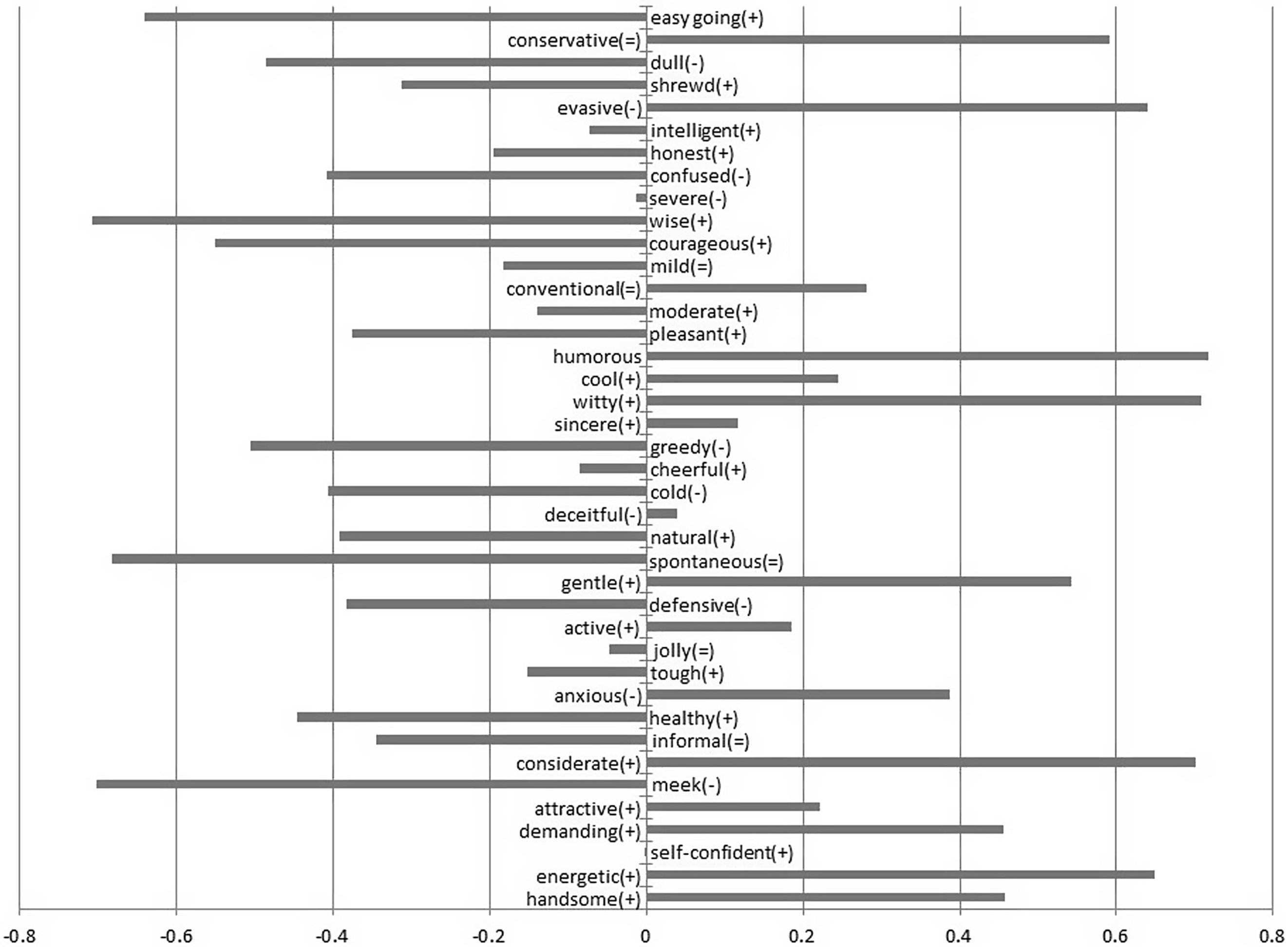

FIG. 2. Trait scores for Romney. Traits arranged from top to bottom in decreasing frequency of volume of discussion. Traits marked + are positive attributes, − are negative attributes, and = are neutral attributes. Equation 3 is used to calculate scores along the *x* axis.

identified correctly: for example, "OMGosh! How can she say Obama is strong in foreign policy?! He isn't." Second, the trait found in the tweet was not written in the context of the politician but connected with something else in the tweet. The false negative errors occurred mostly due to the use of personal pronouns ("he") and words that are synonymous to the politician names such as "Prez" or "the President" for Obama and the "GOP nominee" or "Republican candidate" for Romney, which were missed by the templates.

Results

Personality Traits: Romney Versus Obama

Figures 2 and 3 summarize perceptions of traits of Romney and Obama, respectively. Their *x* axes provide scores as calculated by Equation 3. The plots are limited to the 40 most frequently discussed traits (of the 110) for each politician. We selected traits for which we had at least 1,500 tweets. The number of tweets retrieved for these 40 range from 1,554 to 21,518 for Romney and 2,293 to 35,652 for Obama. The traits are plotted with the most frequently discussed adjective at the top and the adjective discussed the least at the bottom of the *y* axis. This volume of discussion includes tweets mentioning the trait (both its synonyms and its antonyms) as being present or as absent. *Easy going* is the most discussed trait for both.

Thirty-three traits appear in both top-40 lists. Despite this high overlap, key differences exist in the direction and magnitude of these traits across individuals. Table 6 summarizes these differences. The table lists the individual scores and their difference for the 33 shared traits in the top 40. The table also indicates, in the "Advantage" column, which candidate has the advantage on a trait. To have an observable advantage the individual has to lead by at least +0.05 (a somewhat arbitrary cutoff at this point) on a positive trait or lag by at least −0.05 on a negative trait. Neutral traits are treated the same as positive for this analysis: for example, *jolly, mild, informal, spontaneous*, and *conservative*.

Five traits have no observable differences. For the 28 where there is an advantage for one or the other, Obama has

DOI: 10.1002/asi

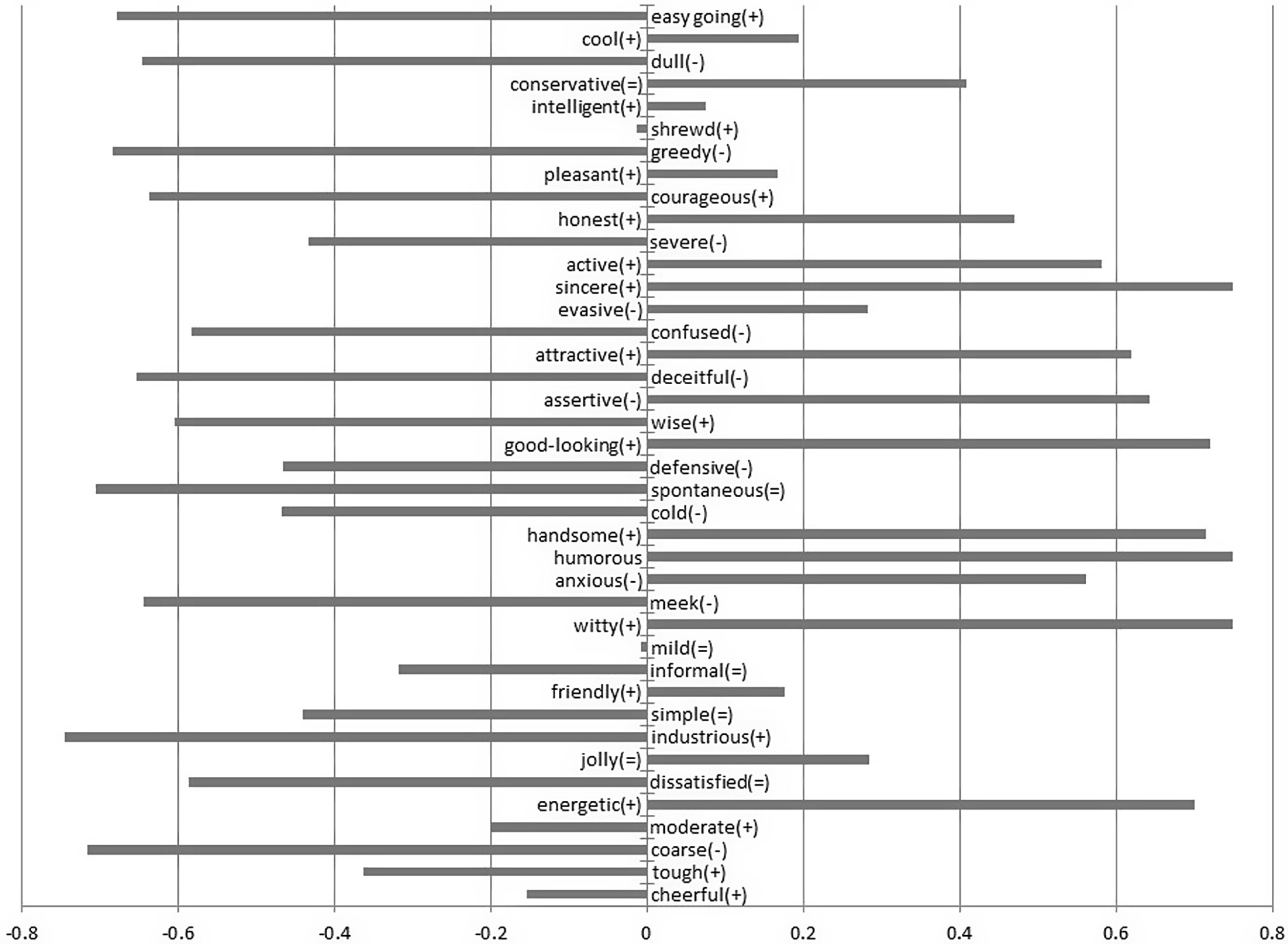

FIG. 3. Trait scores for Obama. Traits arranged from top to bottom in decreasing frequency of volume of discussion. Traits marked + are positive attributes, − are negative attributes, and = are neutral attributes. Equation 3 is used to calculate scores along the *x* axis.

the advantage in 20 of the 28 traits. If we ignore traits marked as neutral (designated by = in column 1), Obama has the advantage in 18 of 25 traits. Some of the notable strong points for Obama are that he is perceived as more *honest, sincere, pleasant, attractive, intelligent* and even less *severe* or *greedy* or *confused* compared to Romney. Romney has the advantage in that he is perceived as more *conservative* and *cool* and less *anxious* than Obama. Both individuals are viewed as *witty, humorous*. Both are equally viewed as lacking in *informality* and *spontaneity*. And neither is *easy going*. Obama rates higher than Romney on more positive traits and lower than Romney on more negative traits. When we repeat this analysis on all 110 traits we find that Obama has the advantage in 51 traits and Romney has the advantage in 39 traits. Thus we see the same trend with the full set of Obama having the overall advantage.

Personality Dimensions: Obama Versus Romney

The top-40 trait list gives a picture of the most frequently mentioned characteristics. Next, we analyze the 14 personality dimensions identified by Simonton (1986). The characterization of the two candidates in terms of these personality measures is summarized in Figure 4. The *x* axis provides scores as calculated by Equation 5.

Figure 4 shows, for example, that for the Friendliness dimension, which can be viewed as a positive quality, Obama and Romney are both perceived as friendly with Obama faring better than Romney. On the other hand, for Machiavellianism, which may be viewed as either a positive or a negative personality dimension, Obama is mostly perceived as not possessing that quality while Romney is perceived as having that quality to some extent. This is the dimension on which perceptions of the two candidates differ the most.

Table 7 summarizes the key differences in direction and magnitude for the 14 personality dimensions. The table is ordered by the last column. Besides Friendliness, the dimensions where both individuals are perceived in the same direction are Poise and Polish, Forcefulness, Wit, Physical Attractiveness, Pettiness, Tidiness, and Conservatism. Out of these eight dimensions, Physical Attractiveness garners

TABLE 6. Differences in top 33 traits shared.

Rating	Trait	Romney	Obama	Obama-Romney	Advantage
+	Honest	−0.196	0.469	0.665	Obama
+	Sincere	0.116	0.748	0.632	Obama
+	Pleasant	−0.375	0.167	0.542	Obama
+	Attractive	0.221	0.620	0.399	Obama
+	Active	0.185	0.581	0.396	Obama
=	Jolly	−0.047	0.284	0.331	Obama
+	Shrewd	−0.312	−0.013	0.299	Obama
+	Handsome	0.457	0.715	0.257	Obama
=	Mild	−0.182	−0.009	0.173	Obama
−	Anxious	0.388	0.561	0.173	Romney
+	Intelligent	−0.073	0.075	0.147	Obama
+	Wise	−0.706	−0.604	0.102	Obama
−	Meek	−0.702	−0.645	0.058	Romney
+	Energetic	0.650	0.700	0.050	Obama
+	Witty	0.709	0.748	0.040	—
*	Humorous	0.718	0.749	0.031	—
=	Informal	−0.345	−0.317	0.028	—
=	Spontaneous	−0.682	−0.706	−0.025	—
+	Easy going	−0.641	−0.678	−0.037	—
+	Cool	0.244	0.194	−0.051	Romney
+	Moderate	−0.140	−0.200	−0.060	Romney
−	Cold	−0.405	−0.467	−0.062	Obama
+	Cheerful	−0.086	−0.155	−0.069	Romney
−	Defensive	−0.383	−0.466	−0.082	Obama
+	Courageous	−0.550	−0.637	−0.087	Romney
−	Dull	−0.486	−0.647	−0.161	Obama
−	Confused	−0.408	−0.582	−0.175	Obama
−	Greedy	−0.505	−0.684	−0.180	Obama
=	Conservative	0.591	0.408	−0.183	Romney
+	Tough	−0.152	−0.363	−0.210	Romney
−	Evasive	0.640	0.283	−0.358	Obama
−	Severe	−0.013	−0.434	−0.421	Obama
−	Deceitful	0.038	−0.653	−0.692	Obama

Note. In first column, * = trait did not receive a majority decision. In last column, — = difference is less than 0.05.

the biggest difference in magnitude (+0.31) in favor of Obama. The two candidates are perceived as closest on Conservatism. Both are perceived as moderately conservative, and the difference between them is tiny (+0.04).

Dimensions for which the two candidates are different in direction are Pacificism, Intellectual Brilliance, Moderation, Achievement Drive, Inflexibility, and Machiavellianism. Besides Machiavellianism, the dimension on which the two candidates differ the most is Pacifism. Pacifism, which may be viewed as a positive quality, has a positive score for Obama (+0.29) and a negative score for Romney (−0.06). For Inflexibility, which has a negative tone, Obama is perceived as just barely negative while Romney is perceived as positive (+0.20). Obama is characterized as negative on Achievement Drive (−0.06) while Romney is characterized as positive (+0.12). Further, Obama is characterized as positive (+0.17) on Intellectual Brilliance while Romney is negative (−0.07).

Across the dimensions, Wittiness, Physical Attractiveness, and Conservatism are the most dominant ones for Obama (scores are >0.50 for these dimensions). For Romney, it is also Wit and additionally Forcefulness. Pettiness is the weakest dimension perceived for both candidates with scores <−0.50 for both.

The qualities on which they are seen in the same light are very important for public figures. However, there are enough differences that the Spearman's rho is a modest 0.59 ($p < 0.05$).

Personality Discussions Over Time

Next, temporal trends in discussion volumes for select personality dimensions are examined. Spikes in discussion volume are accounted for by triggering events. Figure 5 shows the temporal trend in discussion volume for the Moderation personality dimension. The largest Romney spike (May 2, 2012) was on the day former Speaker of the House of Representatives Newt Gingrich suspended his candidacy for the Republican presidential nomination and endorsed Mitt Romney and his policies on the path of moderation. Tweets such as "Newt says he's often asked if Romney is conservative enough and his answer is—'Compared to Barack Obama' " and "Newt Gingrich: Mitt Romney is 'conservative enough' " support the Moderation dimension. Conservative is a synonym of the trait *conventional*, which has a positive loading for the Moderation dimension. The biggest Obama spike was on October 16, 2012, the date of the second presidential debate at Hofstra University in Hempstead, New York, where he emphasized his ideas on moderation in various internal and external affairs. Tweets such as "I think Obama is a cool guy but I don't like some of his proposals" refer to the trait *cool*, which has a positive loading for the Moderation dimension. The September 13 Romney spike coincides with his statements that the world needs a "strong America"[5] and that the United States should be able to participate in two conflicts at once. In this instance, both statements generated tweets on the negative end of Moderation, such as "Romney is impulsive enough to get us into World War 3. No foreign policy knowledge" and "Romney is quite impatient in foreign relations. His knee jerk reaction is a clue."

Figure 6 offers a similar temporal graph for Machiavellianism as a personality dimension. There is a spike for Obama on October 16, 2012 because of tweets such as "@BarackObama is reliable. His support of Israel is real and tangible and can be counted."

Figure 7 is a temporal graph for the dimension Intellectual Brilliance. A key incident that generated negative comments on Obama's intelligence was the episode of the open mike conversation with Medvedev on March 26, 2012 about giving the United States some space on the nuclear issue.[6] An example tweet: "Obama is clearly stupid to be POTUS. The open mic Medved remark is just confirmation. #tcot."

[5]http://blogs.wsj.com/washwire/2012/09/13/romney-says-world-needs-strong-america/

[6]http://www.guardian.co.uk/world/2012/mar/26/obama-medvedev-space-nuclear

DOI: 10.1002/asi

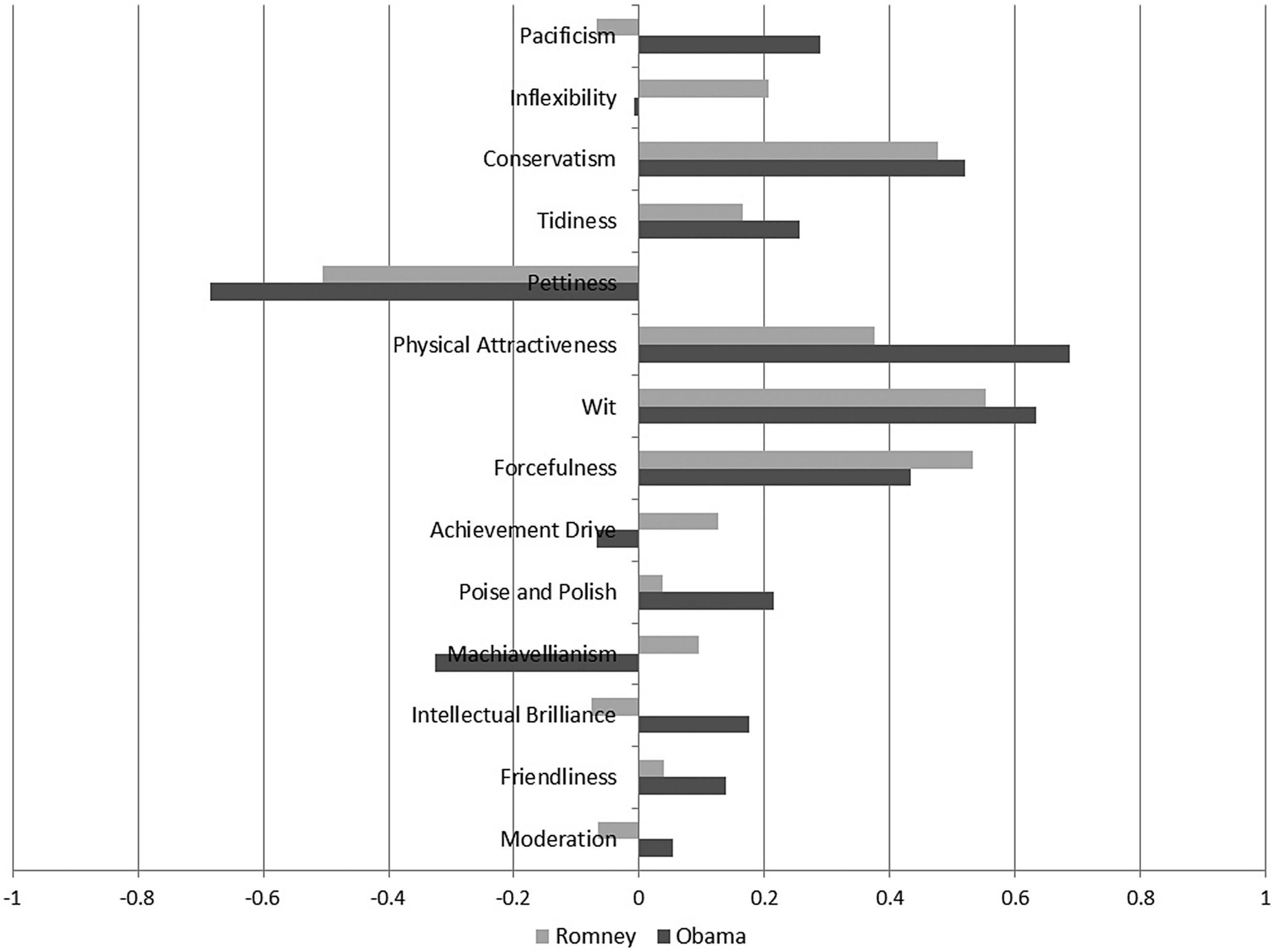

FIG. 4. Scores for Obama and Romney for the 14 personality dimensions.

TABLE 7. Differences in 14 personality dimensions.

Dimension	Romney	Obama	Obama-Romney
Pacifism	–0.068	0.289	0.357
Physical Attractiveness	0.375	0.687	0.312
Intellectual Brilliance	–0.076	0.176	0.252
Poise and Polish	0.037	0.215	0.178
Moderation	–0.066	0.053	0.119
Friendliness	0.04	0.137	0.097
Tidiness	0.165	0.256	0.091
Wit	0.552	0.634	0.082
Conservatism	0.476	0.521	0.045
Forcefulness	0.532	0.433	–0.099
Pettiness	–0.505	–0.684	–0.179
Achievement Drive	0.125	–0.069	–0.194
Inflexibility	0.206	–0.009	–0.215
Machiavellianism	0.095	–0.325	–0.42

On the same day, the tweet “Obama is stupid just like you black slaves in my mentions, what has Obama done for you???” received thousands of retweets. On February 28, 2012, the spike corresponds to the many complimentary posts made in the context of the auto bailouts.[7] Example tweets are: “Overheard on campus: “Barack Obama is a smart guy—he knows what’s going on” and “Obama is smart: rebounding automakers adding jobs.”

Figure 8 shows the temporal plot for discussion level on Friendliness as a personality dimension. The spike on July 27, 2012 for Romney was because of his gaffe in London when he made some negative comments about the Summer Olympics being hosted there.[8] These comments were viewed as detrimental to the friendly relationship between the United Kingdom and the United States and generated posts on the negative side of Friendliness. Examples are: “Even England thinks Romney is rude. #Obama2012” and “WE are not rude. ROMNEY is rude. and the brits are not famous for their cultural sensitive, really, either.” The spike in the Friendliness dimension for Obama around September 20, 2012 is due to the many retweets of the tweet “Ehud Barak: Obama ‘is friendly to Israel, especially in

[7]http://cnnmon.ie/wcm2ig

[8]http://www.cnn.com/2012/07/27/politics/romney-london-troubles

DOI: 10.1002/asi

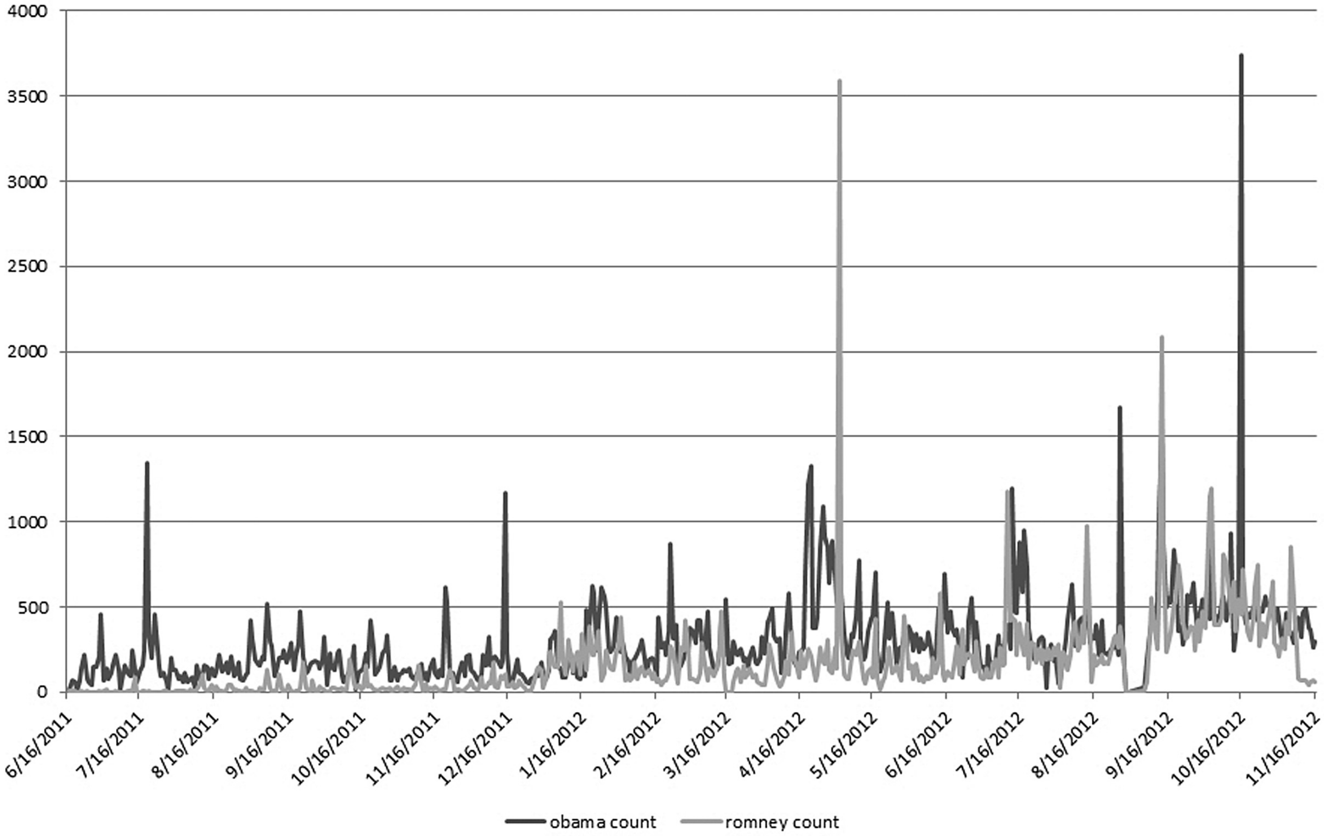

FIG. 5. Discussion trend for Moderation.

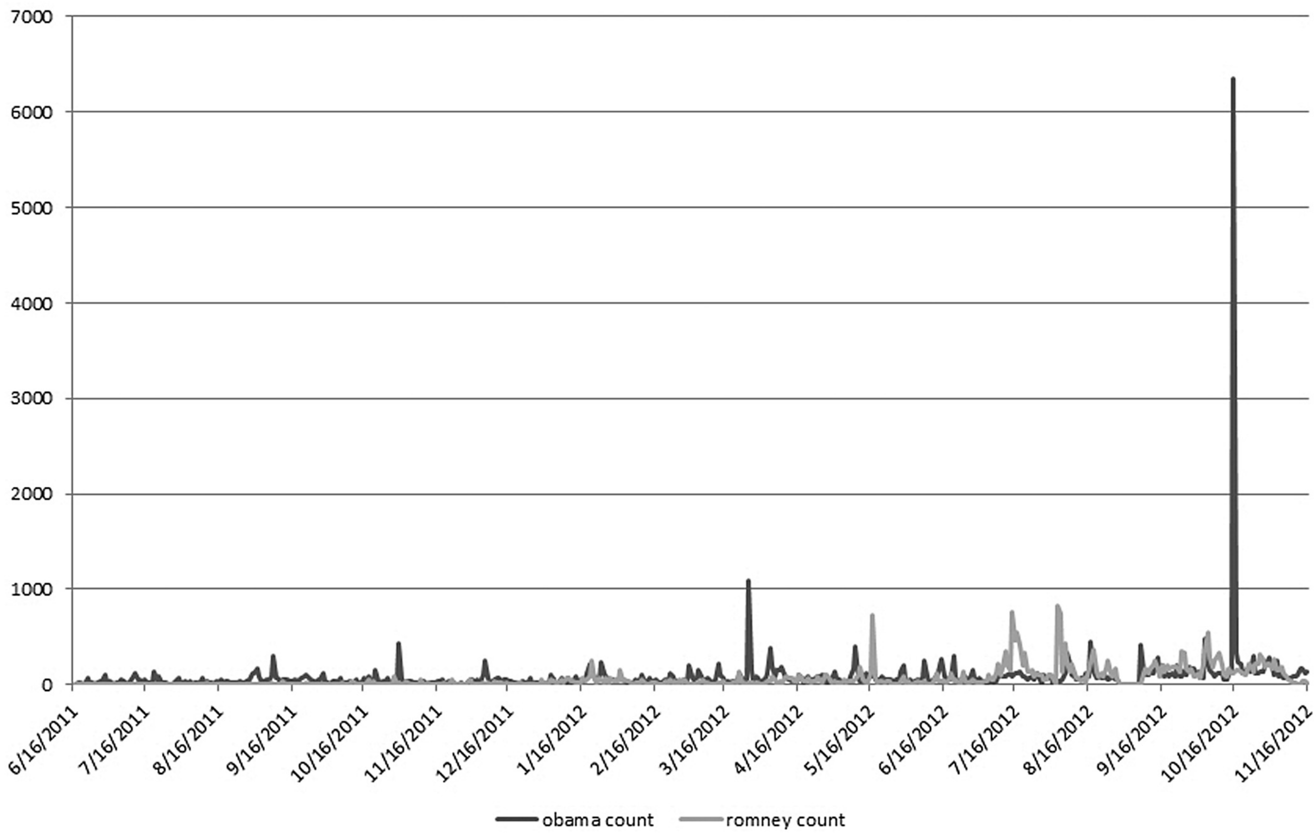

FIG. 6. Discussion trend for Machiavellianism.

DOI: 10.1002/asi

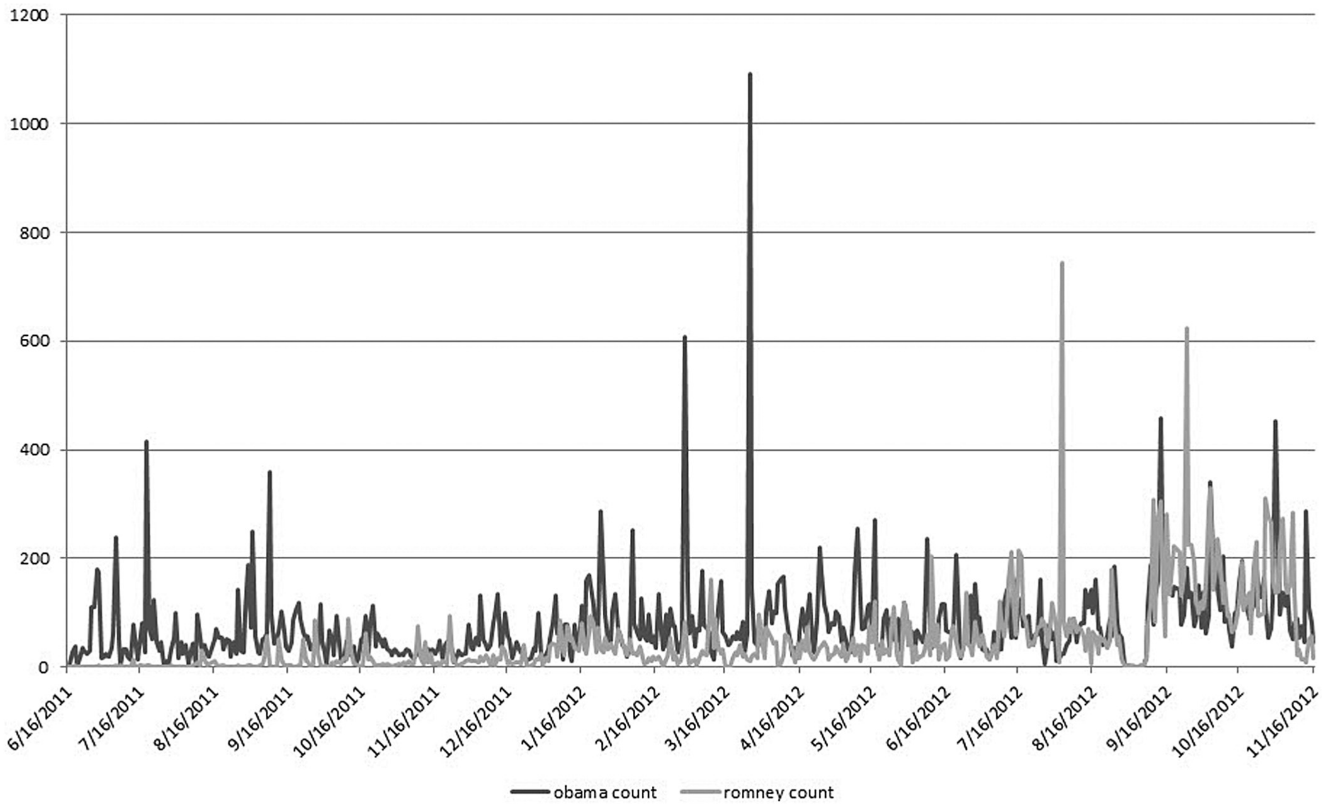

FIG. 7. Discussion trend for Intellectual Brilliance.

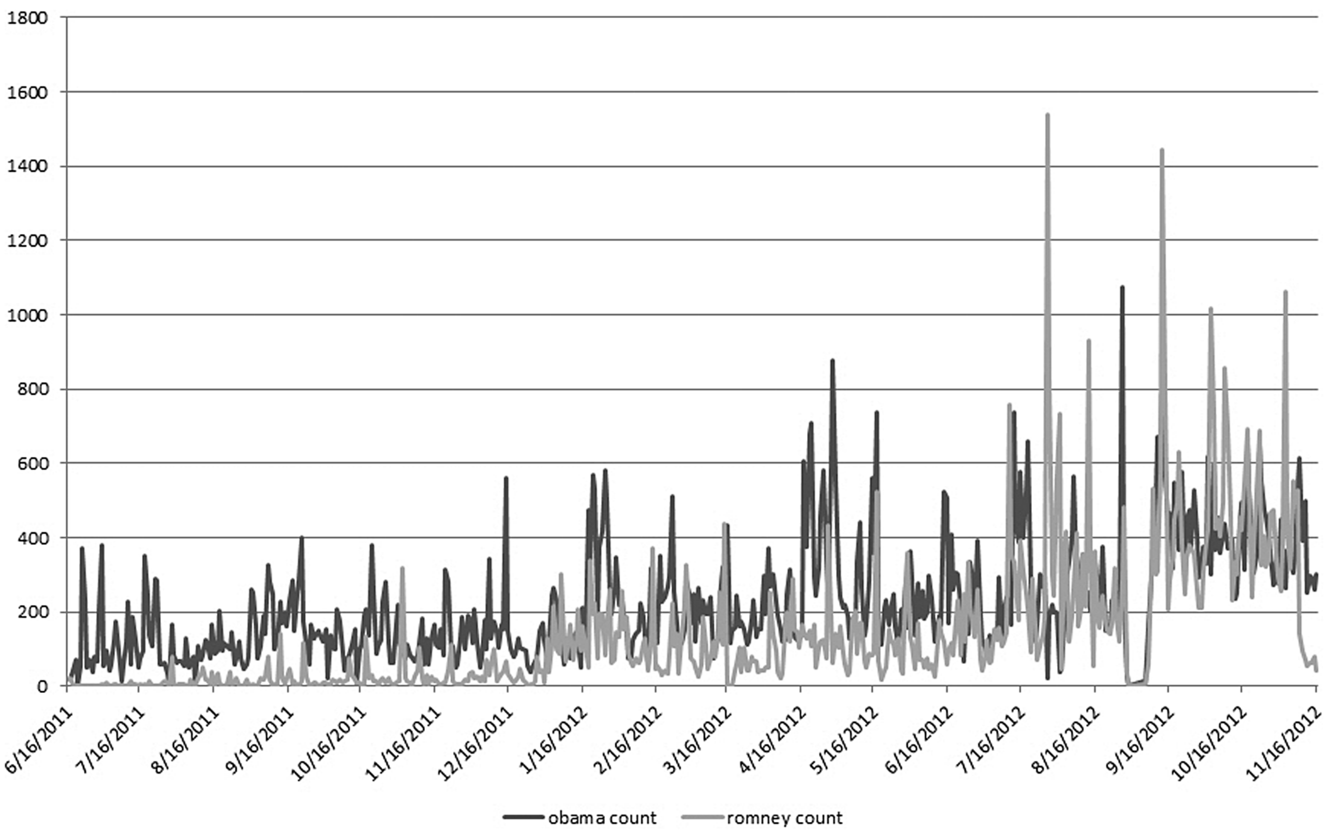

FIG. 8. Discussion trend for Friendliness.

TABLE 8. Correlations between the Gallup poll measure of support and the dimensions of personality.

Dimensions	Obama's positive loadings	Obama's negative loadings	Romney's positive loadings	Romney's negative loadings
Conservatism	–0.484	—	0.401	—
Moderation	–0.045	–0.462	0.378	0.293
Achievement Drive	0.413	–0.400	0.410	0.275
Pacifism	–0.216	–0.494	0.215	0.529
Forcefulness	–0.441	—	0.123	—
Friendliness	–0.017	–0.291	0.352	0.435
Intellectual Brilliance	–0.122	–0.361	–0.020	0.340
Machiavellianism	0.168	0.065	0.073	0.246
Poise and Polish	–0.066	0.193	0.206	0.129
Wit	0.184	—	0.261	—
Physical Attractiveness	0.182	—	0.158	—
Pettiness	0.221	—	–0.118	—
Tidiness	0.062	–0.494	–0.140	0.529
Inflexibility	–0.019	—	0.058	—

Note. The "—" indicates no tweets were found.

security-related issues,' " which correspond to Ehud Barak's interview with CNN when he commented about Obama's friendly Middle East policy.[9]

Overall, these spikes in discussion volume for the personality dimensions analyzed are meaningful and consistent with key events and to that extent validate our methods.

Personality Perceptions and the 2012 Elections

Personality is the foundation of much political advertising (Boynton, 1996; Boynton & Nelson, 1997). For campaign managers, it is a major focus. Does this research of perceptions of the two candidates find that the perceptions are important? If survey research and the Twitter messages reveal similar perceptions of personality, and changes in the distribution of personality-related tweets through time are parallel to changes in support for individual candidates and if the distribution of perceptions of favorable traits match the outcome of the election that would be evidence that it is important to understand perceptions of the personality of candidates.

As far back as 2004, Pew Foundation surveys asked respondents for the one word that best described a candidate such as Bush, Obama, and McCain. Several (but not all) answers were personality traits. In two surveys conducted toward the end of August 2012 (Pew Research Center for the People & the Press, 2012a, 2012b) among the 80 and 78 distinct, top-ranking words generated for Obama and Romney, respectively, we found eight traits for which we had Twitter counts: *bossy*, *deceitful*, *distrustful*, *good-natured*, *honest*, *intelligent*, and *sincere* and *friendly*. With the exception of *friendly*, the candidate with the advantage was the same in both the Pew data and the Twitter data. Pew found Obama friendlier whereas Twitter selected Romney. This comparison is no doubt limited because of the significant differences, for example, between the samples surveyed by Pew and the Twitter users involved in our study.

Elections are long-term events in the United States. The election of 2012 began for Republicans in February 2011. That gave potential voters 21 months to size up the candidates, to form an assessment of who they are and whom they wanted to support. Both perceptions of the candidates and support change slowly. It is necessary to look at changes over periods of time to determine how the two are changing. The time period with the greatest difference between the candidates in the Gallup poll was August 30 to October 7 of 2012 (Gallup, 2013). This was a particularly critical period of the campaign. The Gallup poll measure of support is calculated as a seven-day rolling average. For this period, counts of mentions related to the 14 personality dimensions, both positive and negative, were also calculated using a 7 day rolling average. Correlations were then run between the Gallup poll measure of support and the dimensions of personality using both the positive and negative attributions of the dimension. The correlations are summarized in Table 8.

The personality dimension that offers the most interesting analysis is Conservatism. In political research conservatism is interpreted in policy terms; political parties exhibit differences on policies like in health care. Potential Obama voters did not want a candidate who is conservative, and potential Romney voters prized a candidate who is conservative. Our Twitter study is of personality perceptions and is not explicitly tied to policy. Despite these differences, we find some meaningful correlations between the Twitter observations and Gallup poll data. When mentions of Obama being conservative increased his Gallup support also went down, –0.48. But when mentions of Romney as conservative increased his Gallup support also went up, +0.40. Movements in the other direction for each individual in Twitter were not accompanied by observable changes in Gallup support.

Moderation has something of the same relationship with Gallup support. When mentions of Obama not being

[9]http://piersmorgan.blogs.cnn.com/2011/09/20/ehud-barak-obama-is-friendly-to-israel-especially-in-security-related-issues/

DOI: 10.1002/asi

moderate increased this was accompanied by support going down, –0.46. When mentions of Romney's moderation increased this was accompanied by support also going up though not as much, +0.38. Moderation and Conservatism are two personality dimensions that undergird the policy differences in voter perceptions of the candidates.

For both Obama and Romney, increases in mentioning them as having high achievement drive were accompanied by increases in Gallup poll support, 0.41. Additionally, for Obama, when mentions of lack of achievement drive increased this was accompanied by a fall in support, –0.40.

Pacifism and forcefulness are personality traits that are rather parallel. The Pacifism dimension ranges from pacifism to aggression. We see that when mentions of Obama as aggressive increase, this is accompanied by a decline in Gallup support, –0.49. Also when mentions of Obama as forceful increase, support again declines, –0.44. For Romney, increases in mentions of not pacific, or aggressiveness, correlated with increased support. Perceived aggression in Twitter correlates with Gallup support for Romney, but not for Obama.

Conservatism, Moderation, and Pacifism can be related to policy stance and we observe clear differences between perceptions of the two candidates along these dimensions in Twitter. The remaining dimensions indicate low correlations with Gallup support or are less interesting.

Obama won and Romney lost. Are our results on perceptions of them at the trait level consistent with this outcome? The 33 traits most frequently mentioned in Twitter divide into 20 on which Obama was perceived more positively than Romney, eight on which Romney was perceived more positively than Obama, and five for which there was no difference. The perceptions of personality traits have a winner who happens to be the same as the winner of the election. And the difference between the two sets of Twitter perceptions is not small.

Conclusions

Public opinions on the personalities of politicians have an impact on their success in politics broadly and also specifically in events such as elections. Twitter with its many spontaneous affect expressions offers a basis for the study of perceptions of personality. We propose this as an important goal that is clearly relevant to more detailed monitoring of public opinion before and during elections. The high-precision, template-driven approach for finding the right tweets and extracting trait-specific opinions provides a methodological tool for the research. The method is strong and may be applied easily to other politicians and other electoral contexts as well as to individuals in other public spheres of life (e.g., actors, corporate leaders, philanthropists, and writers). The method also forms a distinct complement to the mainstream sentiment analysis methods using lexicon-based frequency counts and classifiers. Note that although the ACL lexicon of adjectives is used, these are enmeshed into retrieval templates designed to capture common ways of expressing opinion on personality traits. Obama is perceived by Twitter users as moderate, intellectually brilliant, and a pacifist. Romney is perceived as having high achievement drive and being Machiavellian as well as inflexible.

The opinions are sometimes remarkably different across candidates, such as on honesty and anxiety. When aggregated into the personality dimensions, strong differences are found in, for example, Moderation and Pacifism. Finally, when studying temporal trends in the volume of discussion related to the personality dimensions, the observations are consistent with expectations around key events. For example, for the Moderation dimension Romney's trend shows a spike around the Republican date of nomination. This happened because of an increase in tweets on Romney being considerate, tactful, unassuming, etc. Overall, Obama, the winner of the 2012 election, had the advantage over Romney in what Twitter said about their personalities. Comparisons with Gallup support polls and with Pew data also provide meaningful inferences. These indicate that campaign managers may find it advantageous to monitor perceptions on Twitter in the months preceding an election.

There are several ways in which to strengthen and extend the current work. One is augmenting the trait list by including traits such as fairness, balance, and generosity. The ACL lexicon is a great starting point, widely acknowledged in psychology and political science, but it can be extended to capture more aspects about personality. Second, the recall of the methods should be improved. Although performance is already good given the noisiness of social media, we will aim for improvements. Another extension is exploring the social network behind the tweeters expressing a certain class of opinion (e.g., tweets indicating that Obama is honest or Romney is forceful). Are there identifiable communities behind certain classes of opinions? Do certain types of opinions on personality traits spread faster than other types? These and other follow-up questions will be the basis of our continuing research.

References

Al-Khalifa, H.S. (2011). Exploring political activities in the Saudi Twitterverse. In Proceedings of the 13th International Conference on Information Integration and Web-based Applications and Services (iiWAS ′11) (pp. 363–366). New York, NY: ACM.

Allport, G.W., & Odbert, H.S. (1936). Trait-names: A psycho-lexical study. Psychological Monographs, 47(1), i–171.

Boutet, A., Kim, H., & Yoneki, E. (2013). What's in Twitter: I know what parties are popular and who you are supporting now! Social Network Analysis and Mining, 3(4), 1379–1391.

Boynton, G.R. (1996). The art of campaign advertising. Chatham, NJ: Chatham House.

Boynton, G.R., & Nelson, J.S. (1997). Rhetorics of political advertising. Urbana, IL: University of Illinois Press.

Bravo-Marquez, F., Gayo-Avello, D., Mendoza, M., & Poblete, B. (2012). Opinion dynamics of elections in Twitter. In Proceedings of the 2012 Eighth Latin American Web Congress (LA-WEB '12) (pp. 32–39). Washington, DC: IEEE Computer Society.

Campbell, A., Converse, P.E., Miller, W.E., & Stokes, D.E. (1960). The American voter. Unabridged edition. Chicago, IL: University of Chicago Press.

DOI: 10.1002/asi

Caprara, G.V., & Cervone, D. (2000). Personality: Determinants, dynamics, and potentials. New York: Cambridge University Press.
Chung, J.E., & Mustafaraj, E. (2011). Can collective sentiment expressed on Twitter predict political elections? In Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence (pp. 1770–1771). San Francisco, CA: AAAI Press.
Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., & Flammini, A. (2011a). Political polarization on Twitter. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (pp. 89–96). Barcelona, Spain: AAAI Press.
Conover, M.D., Gonçalves, B., Ratkiewicz, J., Flammini, A., & Menczer, F. (2011b). Predicting the political alignment of twitter users. In Privacy, Security, Risk and Trust (PASSAT), 2011 IEEE Third International Conference on Social Computing (SocialCom) (pp. 192–199). Washington, DC: IEEE Computer Society.
Contractor, D., & Faruquie, T.A. (2013). Understanding election candidate approval ratings using social media data. In Proceedings of the 22nd International Conference on World Wide Web Companion (pp. 189–190). Geneva, Switzerland: International World Wide Web Conferences Steering Committee.
Costantini, E., & Craik, K.H. (1980). Personality and politicians: California party leaders, 1960–1976. Journal of Personality and Social Psychology, 38(4), 641.
Diakopoulos, N.A., & Shamma, D.A. (2010). Characterizing debate performance via aggregated Twitter sentiment. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1195–1198). New York: ACM.
Diaz-Aviles, E., Orellana-Rodriguez, C., & Nejdl, W. (2012). Taking the pulse of political emotions in Latin America based on social web streams. In Proceedings of the 2012 Eighth Latin American Web Congress (LA-WEB ‘12) (pp. 40–47). Washington, DC: IEEE Computer Society.
Eveland, W.P., & Hutchens, M.J. (2013). The role of conversation in developing accurate political perceptions: A multilevel social network approach. Human Communication Research, 39(4), 422–444.
Gallagher, M.E., & Allen, S.H. (2014). Presidential personality: Not just a nuisance. Foreign Policy Analysis, 10(1), 1–21.
Gallup (2013). Election 2012 Registered Voters Trial Heat: Obama versus Romney. Retrieved from http://www.gallup.com/poll/150743/obama-romney.aspx
Gaurav, M., Srivastava, A., Kumar, A., & Miller, S. (2013). Leveraging candidate popularity on Twitter to predict election outcome. In Proceedings of the Seventh Workshop on Social Network Mining and Analysis (Article 7). New York: ACM.
Gayo-Avello, D. (2012). No, you cannot predict elections with Twitter. Internet Computing Society, IEEE, 16(6), 91–94.
Gayo-Avello, D., Metaxas, P.T., & Mustafaraj, E. (2011). Limits of electoral predictions using Twitter. In Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (pp. 490–493). Barcelona, Spain: AAAI Press.
George, A.L., & George, J.L. (1998). Presidential personality and performance. Boulder, CO: Westview Press.
Golbeck, J., & Hansen, D. (2011). Computing political preference among Twitter followers. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1105–1108). New York: ACM.
Gough, H.G., & Heilbrun, A.B. (1983). The adjective check list manual. Mountain View, CA: Consulting Psychologists Press.
Hadgu, A.T., Garimella, K., & Weber, I. (2013). Political hashtag hijacking in the U.S. In Proceedings of the 22nd International Conference on World Wide Web Companion (pp. 55–56). Geneva, Switzerland: International World Wide Web Conferences Steering Committee.
Hanna, A., Wells, C., Maurer, P., Friedland, L., Shah, D., & Matthes, J. (2013). Partisan alignments and political polarization online: A computational approach to understanding the French and U.S. presidential elections. In Proceedings of the Second Workshop on Politics, Elections and Data (pp. 15–22). New York: ACM.
Hemphill, L., Otterbacher, J., & Shapiro, M. (2013). What’s Congress doing on Twitter? In Proceedings of the 2013 Conference on Computer Supported Cooperative Work (pp. 877–886). New York: ACM.
Hong, S., & Nadler, D. (2011). Does the early bird move the polls? The use of the social media tool “Twitter” by U.S. politicians and its impact on public opinion. In Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times (pp. 182–186). New York: ACM.
Huberty, M.E. (2013). Multi-cycle forecasting of congressional elections with social media. In Proceedings of the Second Workshop on Politics, Elections and Data (pp. 23–30). New York: ACM.
Jansen, B.J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. Journal of the American Society for Information Science and Technology, 60(11), 2169–2188.
Jungherr, A. (2013). Tweets and votes, a special relationship: The 2009 federal election in Germany. In Proceedings of the Second Workshop on Politics, Elections and Data (pp. 5–14). New York: ACM.
Kim, J., & Yoo, J. (2012). Role of sentiment in message propagation: Reply versus retweet behavior in political communication. In Social Informatics (SocialInformatics), 2012 International Conference (pp. 131–136). Washington, DC: IEEE Computer Society.
Kinder, D.R. (1978). Political person perception: The asymmetrical influence of sentiment and choice on perceptions of presidential candidates. Journal of Personality and Social Psychology, 36(8), 859–871.
Marchetti-Bowick, M., & Chambers, N. (2012). Learning for microblogs with distant supervision: Political forecasting with Twitter. In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (pp. 603–612). Stroudsburg, PA: Association for Computational Linguistics.
Mejova, Y., & Srinivasan, P. (2012). Political speech in social media streams: YouTube comments and Twitter posts. In Proceedings of the Third Annual ACM Web Science Conference (pp. 205–208). New York: ACM.
Mejova, Y., Srinivasan, P., & Boynton, B. (2013). GOP primary season on Twitter: Popular political sentiment in social media. In Proceedings of the Sixth ACM International Conference on Web Search and Data Mining (pp. 517–526). New York: ACM.
Melville, P., Gryc, W., & Lawrence, R.D. (2009). Sentiment analysis of blogs by combining lexical knowledge with text classification. In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1275–1284). New York: ACM.
Metaxas, P.T., Mustafaraj, E., & Gayo-Avello, D. (2011). How (not) to predict elections. In Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom) (pp. 165–171). Washington, DC: IEEE Computer Society.
Mohammad, S.M., Kiritchenko, S., & Martin, J. (2013). Identifying purpose behind electoral tweets. In Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining (Article 1). New York: ACM.
Monti, C., Zignani, M., Rozza, A., Arvidsson, A., Zappella, G., & Colleoni, E. (2013). Modelling political disaffection from Twitter data. In Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining (Article 3). New York: ACM.
Mustafaraj, E., Finn, S., Whitlock, C., & Metaxas, P.T. (2011). Vocal minority versus silent majority: Discovering the opinions of the long tail. In Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom) (pp. 103–110). Washington, DC: IEEE Computer Society.
Nooralahzadeh, F., Arunachalam, V., & Chiru, C.G. (2013). 2012 presidential elections on Twitter—An analysis of how the U.S. and French election were reflected in tweets. In Control Systems and Computer Science (CSCS), 2013 19th International Conference (pp. 240–246). Washington, DC: IEEE Computer Society.
O’Reilly, C.A., Chatman, J., & Caldwell, D.F. (1991). People and organizational culture: A profile comparison approach to assessing person-organization fit. Academy of Management Journal, 34(3), 487–516.
Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC′10) (pp. 1320–1326). Valletta, Malta: European Language Resources Association.

DOI: 10.1002/asi

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1–2), 1–135.
Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. In Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing (Vol. 10, pp. 79–86). Stroudsburg, PA: Association for Computational Linguistics.
Pew Research Center for the People & the Press. (2012a). Obama in a word: From “inexperienced” and “change” to “good” and “trying.” Retrieved from http://www.people-press.org/2012/09/05/obama-in-a-word-from-inexperienced-and-change-to-good-and-trying/
Pew Research Center for the People & the Press. (2012b). Romney in a word: “Honest”, “businessman”, “rich.” Retrieved from http://www.people-press.org/2012/08/29/romney-in-a-word-honest-businessman-rich/
Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., & Menczer, F. (2011). Truthy: Mapping the spread of astroturf in microblog streams. In Proceedings of the 20th International Conference Companion on World Wide Web (pp. 249–252). New York: ACM.
Sang, E.T.K., & Bos, J. (2012). Predicting the 2011 Dutch senate election results with Twitter. In Proceedings of the Workshop on Semantic Analysis in Social Media (pp. 53–60). Stroudsburg, PA: Association for Computational Linguistics.
Simonton, D.K. (1986). Presidential personality: Biographical use of the Gough Adjective Check List. Journal of Personality and Social Psychology, 51(1), 149–160.
Skoric, M., Poor, N., Achananuparp, P., Lim, E.P., & Jiang, J. (2012). Tweets and votes: A study of the 2011 Singapore general election. In HICSS ’12: Proceedings of the 2012 45th Hawaii International Conference on System Sciences (pp. 2583–2591). Washington, DC: IEEE Computer Society.
Small, T. (2010). Canadian politics in 140 characters: Party politics in the Twitterverse. Canadian Parliamentary Review, 33(3), 39–45.
Soler, J.M., Cuartero, F., & Roblizo, M. (2012). Twitter as a tool for predicting elections results. In Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012) (pp. 1194–1200). Washington, DC: IEEE Computer Society.
Starbird, K., & Palen, L. (2012). How will the revolution be retweeted? Information diffusion and the 2011 Egyptian uprising. In Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (pp. 7–16). New York: ACM.
Stieglitz, S., & Dang-Xuan, L. (2012). Political communication and influence through Microblogging—An empirical analysis of sentiment in Twitter messages and retweet behavior. In HICSS ’12: Proceedings of the 2012 45th Hawaii International Conference on System Sciences (pp. 3500–3509). Washington, DC: IEEE Computer Society.
Strohman, T., Metzler, D., Turtle, H., & Croft, W.B. (2005). Indri: A language model-based search engine for complex queries . CIIR Technical Report. Amherst, MA.
Tausczik, Y.R., & Pennebaker, J.W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. Journal of Language and Social Psychology, 29(1), 24–54.
Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. Journal of the American Society for Information Science and Technology, 63(1), 163–173.
Tumasjan, A., Sprenger, T.O., Sandner, P.G., & Welpe, I.M. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. In Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media (pp. 178–185). Washington, DC: AAAI Press.
Tupes, E.C., & Christal, R.E. (1992). Recurrent personality factors based on trait ratings. Journal of Personality, 60(2), 225–251.
Wang, H., Can, D., Kazemzadeh, A., Bar, F., & Narayanan, S. (2012). A system for real-time Twitter sentiment analysis of 2012 U.S. presidential election cycle. In Proceedings of the ACL 2012, System Demonstrations (pp. 115–120). Stroudsburg, PA: Association for Computational Linguistics.
Wegrzyn-Wolska, K., & Bougueroua, L. (2012). Tweets mining for French presidential election. In Computational Aspects of Social Networks (CASoN), 2012 Fourth International Conference (pp. 138–143). Washington, DC: IEEE Computer Society.
Wei, Z., He, Y., Gao, W., Li, B., Zhou, L., & Wong, K.F. (2013). Mainstream media behavior analysis on Twitter: A case study on UK general election. In Proceedings of the 24th ACM Conference on Hypertext and Social Media (pp. 174–178). New York: ACM.
Wu, Y., Wong, J., Deng, Y., & Chang, K. (2011). An exploration of social media in public opinion convergence: Elaboration likelihood and semantic networks on political events. In Dependable, Autonomic and Secure Computing (DASC), 2011 IEEE Ninth International Conference (pp. 903–910). Washington, DC: IEEE Computer Society.
Younus, A., Qureshi, M.A., Asar, F.F., Azam, M., Saeed, M., & Touheed, N. (2011). What do the average twitterers say: A Twitter model for public opinion analysis in the face of major political events. In Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference (pp. 618–623). Washington, DC: IEEE Computer Society.

DOI: 10.1002/asi

Appendix

A summary of recent Twitter-based political research.

Author and year	Targeted subject	Problem addressed	Method	Sentiment analysis type	Predict election?	Select conclusions
Monti et al., 2013	Italian politicians in 2012	Political disaffection	Use classifiers to determine politically disaffected tweets, and produced time-series	Negative sentiment	Yes	Important political news of Italian newspapers is often correlated with the highest peaks of the produced time-series.
Nooralahzadeh et al., 2013	Obama, Romney, Sarkozy, Hollande in 2012	Political sentiment	Score the sentiment of tweets using lexicon and did time-series analysis	Positive, negative	No	The sentiment of elections reflected the real-time results.
Hanna et al., 2013	Obama, Biden, Ryan, Romney, Hollande, Sarkozy, Le Pen, Bayrou in 2012	Twitter discussion around elections	Counts of candidate mentions, and hashtag usage in retweet	N/A	No	More stark political polarization in the French case, while less partisan division in U.S. case
Hemphill et al., 2013	380 members of U.S. Congress from 2011 to 2012	Explore Congress' Twitter activities	Use classifier to code tweets into 6 classes(e.g., Narrating, Positioning, Thanking, etc.)	N/A	No	Officials frequently use Twitter to advertise their political positions and to provide information, but rarely to request political action from their constituents or to recognize the good work of others.
Mejova et al., 2013	7 U.S .Republicans from 2011 to 2012	Political sentiment	Use classifiers to determine if tweets are for, against, or neutral to the politician	For, against, neutral	No	Twitter political chatter is not indicative of national political polls.
Hadgu et al., 2013	25 U.S. politicians, including Obama	Hashtag usage	Discover sudden changes of hashtag in leaning, and analyze the change points	N/A	No	The sudden usage change of hashtag corresponds to activity by "hashtag hijackers"
Contractor & Faruquie, 2013	Obama, Romney in U.S. presidential elections, 2012	Predict approval ratings	Use keywords match to identify supporter and regression model to predict the poll rating	N/A	Yes	Use tweets to predict the daily approval ratings of two U.S. presidential candidates.
Huberty, 2013	U.S. Republican or Democratic general election candidate from 2010 to 2012	Predict election	Compare different prediction models	N/A	Yes	Simplistic methods for forecasting elections from Twitter; even when their results are correlated with election outcomes, provide relatively little added benefit.
Wei et al., 2013	4 major UK TV media outlets in 2010	Media bias	Use sentiment analysis to measure media bias	N/A	No	Tweets from Journalists are more likely to be retweeted than those from mainstream media
Jungherr, 2013	German parties, Merkel, Steinmeier in 2009	Hashtag usage	Use number of hashtag mentions, #partname+/− to determine sentiment	Positive, Negative	Yes	During the campaign of 2009 Twitter messages commenting on parties and candidates showed little, if any, systematic relationship with subsequent votes on election day.
Mohammad et al., 2013	Obama, Romney in U.S. presidential elections, 2012	Purpose of public political tweets	Use classifier to determine 11 classes of purpose	11 categories: agree, praise, support, disagree, ridicule, etc.	No	Detect electoral tweets' purposes; number of opposing tweets almost twice number of supporting tweets
Gaurav et al., 2013	Candidates in elections in Venezuela, Paraguay, and Ecuador, 2012–2013	Predict election	Number of mentions of candidates' names and aliases	N/A	Yes	Successful in predicting winner of all three presidential elections in Latin America
Mejova & Srinivasan, 2012	Obama, Perry, Gingrich, Bachmann, Paul in 2011	Data source difference; Twitter &YouTube	Manually annotate sentiment of a subset	Positive, negative, mixed, none	Yes	Neither discussion volume nor sentiment expressed in the two media able to predict the Republican presidential nominee front-runner.
Wang et al., 2012	9 Republicans and Obama in 2012 presidential election	Real-time sentiment analysis	Use classifier to detect sentiment	Positive, negative, neutral, unsure	No	Built a system for real-time Twitter sentiment analysis
Marchetti-Bowick & Chambers, 2012	Obama in 2000	Predict election	Use distantly supervised classifiers	Positive, negative	Yes	Method better correlates with Gallup's Presidential Job Approval polls than lexicon-based work.
Sang & Bos, 2012	12 parties in Dutch 2011 election	Predict election	Use entity counts	N/A	Yes	The seat numbers predicted by the tweets were close to the election results.
Diaz-Aviles et al., 2012	18 presidents in Latin America. 2011–2012	Emotion analysis	Use name counts and Plutchik's emotion model lexicon	Plutchik's 8 emotions	Yes	Linear combination of emotions achieved a good prediction performance.
Stieglitz & Dang-Xuan, 2012	5 German parties in 2011	Retweet analysis	Use LIWC lexicon to detect sentiment, regression to examine retweet rate	Positive, negative	No	Sentiment affects retweet rate
Skoric et al., 2012	7 parties in Singapore 2011 election	Predict election	Use number of name mentions	N/A	Yes	Could be used to predict Singapore election, but performance lower than in Germany and UK.

DOI: 10.1002/asi

Appendix Continued

Author and year	Targeted subject	Problem addressed	Method	Sentiment analysis type	Predict election?	Select conclusions
Bravo-Marquez et al., 2012	Obama–Biden, McCain–Palin in 2008	Predict election	Use time-series model to predict election	Positive, negative	Yes	Time series are not fit for predicting future opinion trends.
Soler et al., 2012	7 parties in Spain, 2011	Predict election	Use number of mentions	N/A	Yes	Twitter references to the parties significantly correlated with the votes
Boutet et al., 2013	3 parties in UK 2010 election	Party characteristics, user Classification	Use label propagation method to identify users' affiliations	Activity, influence, structure/interaction, content, & sentiment	No	Labor is active and influential, Conservative is the most organized to promote their activities. Websites and blogs that each political party supported are clearly different from those that all the other political parties supported.
Wegrzyn-Wolska & Bougueroua, 2012	Candidates in French 2012 election	Sentiment analysis	Use 2 step classifiers to extract polarity	Positive, negative, neutral	Yes	Built a system to detect sentiment of tweets
Gayo-Avello, 2012	N/A	Critique election prediction	N/A	N/A	No	Critique of current works
Al-Khalifa, 2011	Saudi political tweets	Hashtag usage	Built network graph for frequent hashtags	N/A	No	Visualize the shape of political networks for different hashtags
Starbird & Palen, 2012	Egyptian political uprisings in 2011	Information diffusion	Analyze retweet statistics	N/A	No	Nearly 60% of tweets sent with tags referencing the Egyptian protests
Kim & Yoo, 2012	Tweets in UK 2010 political election	Sentiment in information propagation	Use LIWC to detect sentiment	Positive, negative	No	Degree of emotion expressions in Twitter messages can affect the number of replies generated as well as retweet rates
Conover et al., 2011a, 2011b	U.S. congressional midterm elections in 2010	User alignment	Use classifier to predict user's alignment	User's left, right, ambiguous	No	Classifier trained on hashtag metadata yield predictions of political affiliations with 91% accuracy.
Chung & Mustafaraj, 2011	Coakley and Brown in Mass. Senate special election, 2010	Predict election	Use OpinionFinder Lexicon to calculate sentiment score for tweets	Support, oppose, neutral	Yes	Methods are not adequate for prediction
Gayo-Avello et al., 2011	U.S. 2010 congressional elections	Predict election	Use OpinionFinder Lexicon to calculate sentiment score for tweets	Positive, negative, or neutral	Yes	Found no correlation between their predictions and the electoral outcomes.
Mustafaraj et al., 2011	Coakley, Brown (Election for U.S. Senate in Mass. 2010)	User group characteristics	Compare two groups using hashtags, content, retweet stats	N/A	No	The content from "vocal minority" and "silent majority" is significantly different.
Hong & Nadler, 2011	9 U.S. Republicans in 2010	Impact of Twitter on public opinion	Use number of mentions	N/A	Yes	Found little evidence that the political use of Twitter has either a positive or negative impact on public opinion.
Golbeck & Hansen, 2011	Twitter users	User preferences	Computed preference score for followers of each member of Congress on Twitter by averaging ADA scores for all congresspeople he or she follows	N/A	No	Political preferences of media outlets' audiences reflect the liberal/conservative leanings of the media outlets as presented in prior literature.
Ratkiewicz et al., 2011	U.S. midterm elections in 2010	Information diffusion	Use number of mentions, GPOMS to detect sentiment	N/A	No	Developed a system for the real-time analysis of meme diffusion from microblog streams
Wu et al., 2011	Singapore 2011 general election	Opinion convergence	Use SentiStrength to detect sentiment, Elaboration Likelihood model to analyze opinion convergence	Positive, negative, neutral	No	Informative tweets were more effective than affective tweets; winners of election were highly connected to the overlapping statements of informative and affective tweets.
Younus et al., 2011	Tunisian uprising in 2011	Subjectivity analysis	Naïve Bayes classifiers with tweet's features	Subjective, objective	No	Achieved accuracy of 83.3%
Metaxas et al., 2011	N/A	Prediction review	N/A	N/A	No	Previous prediction methods are not better than chance.
Diakopoulos & Shamma, 2010	Obama and McCain in 2008 presidential debate	Sentiment to debate	Using MTurk to rate the sentiment of tweets	Negative, positive, mixed, and other	No	Overall sentiment of debate was negative; Tweeters tended to favor Obama over McCain.
Small, 2010	Canadian parties and leaders	Party Twitter usage	Compare Twitter usage for different parties and leaders	N/A	No	Canadian politicians mostly use Twitter to broadcast official party information
Tumasjan et al., 2010	6 parties in German election, 2009	Predict election	Use LIWC to detect sentiment	Positive, negative, sadness, anxiety, anger, etc.	Yes	Number of messages mentioning a party reflects the election result.

DOI: 10.1002/asi