Journal of Research in Personality 46 (2012) 710–718

Contents lists available at SciVerse ScienceDirect

Journal of Research in Personality

journal homepage: www.elsevier.com/locate/jrp

You are what you tweet: Personality expression and perception on Twitter

Lin Qiu *, Han Lin, Jonathan Ramsay, Fang Yang

Division of Psychology, Nanyang Technological University, Singapore

ARTICLE INFO

Article history:
Available online 8 September 2012

Keywords:
Personality
Microblogs
Twitter
Social media
Linguistic analysis

ABSTRACT

Microblogging services such as Twitter have become increasingly popular in recent years. However, little is known about how personality is manifested and perceived in microblogs. In this study, we measured the Big Five personality traits of 142 participants and collected their tweets over a 1-month period. Extraversion, agreeableness, openness, and neuroticism were associated with specific linguistic markers, suggesting that personality manifests in microblogs. Meanwhile, eight observers rated the participants' personality on the basis of their tweets. Results showed that observers relied on specific linguistic cues when making judgments, and could only judge agreeableness and neuroticism accurately. This study provides new empirical evidence of personality expression in naturalistic settings, and points to the potential of utilizing social media for personality research.

1. Introduction

Twitter, one of the most popular microblogging services, has over 200 million users and produces 110 million tweets (i.e., 140-character text messages) every day (Chiang, 2011). It allows users to write short messages and broadcast them to their subscribers, known as followers, in real time. Compared to traditional blogging, microblogging emphasizes speed and brevity, focusing on things happening "right now" (Oulasvirta, Lehtonen, Kurvinen, & Raento, 2010). Several studies have shown that users mainly use microblogs to describe their daily routines, carry out conversations, report news, and share information (e.g., Java, Song, Finin, & Tseng, 2007; Naaman, Boase, & Lai, 2010), producing a text-based "social awareness stream" consisting of everyday thoughts, feelings, and conversations (Naaman et al., 2010).

Despite the increasing prevalence of microblogging, little is known about its association with personality. Do users express their personality in their microblogs? Can one make accurate personality judgments based on others' microblogs? Answers to these questions will significantly advance our understanding of the relationship between personality and social media. Furthermore, microblogs offer a valuable opportunity to investigate personality expression and perception in a naturalistic setting. Current personality studies often rely on self-report surveys in well-controlled, decontextualized environments (Rozin, 2001). However, as Barker and Wright (1951) advocated, we must study natural behaviors in everyday settings to truly understand what people are like. With millions of users contributing to microblogs every day, microblogging produces a vast written record of people's daily behavior in their natural environment. By studying microblogging, we can enhance our understanding of personality as it is expressed in naturally occurring writing samples. Thus, in this study, we aim to understand the relationship between personality and microblogging.

2. Background and research questions

2.1. Behavioral cues to personality

Previous research has shown that people inadvertently leave personality-related "behavioral residue" (Gosling, Ko, Mannarelli, & Morris, 2002, p. 381) in their physical and virtual environments. Examples of personality expression have been found in daily conversations (Mehl, Gosling, & Pennebaker, 2006), bedrooms and offices (Gosling et al., 2002), Facebook profiles (Back et al., 2010; Gosling, Augustine, Vazire, Holtzman, & Gaddis, 2011), and virtual world activities (Yee, Harris, Jabon, & Bailenson, 2011). Since people frequently use microblogs to record their thoughts and activities, it is reasonable to expect that an individual's microblogs will also contain their personality-related residue.

Meanwhile, research on language use has shown a connection between self-reported personality and writing style (Pennebaker & King, 1999). A software program called Linguistic Inquiry and Word Count (LIWC) has been widely used to identify linguistic patterns associated with personality traits by calculating word frequencies in psychologically meaningful categories, such as pronouns, social terms, and affect terms (Pennebaker, Booth, & Francis, 2007). Extraverts have been found to produce fewer large words (Mehl et al., 2006), less complex writings, and more social

* Corresponding author. Address: Division of Psychology, Nanyang Technological University, HSS-04-15, 14 Nanyang Drive, Singapore 637332, Singapore. Fax: +65 6795 5797.

E-mail address: linqiu@ntu.edu.sg (L. Qiu).

0092-6566/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jrp.2012.08.008

and positive emotional words than introverts (Pennebaker & King, 1999). Neuroticism has been found to be associated with the use of more anxiety words (Golbeck, Robles, & Turner, 2011), while agreeable individuals employ more positive emotion words and first person plural pronouns (Yarkoni, 2010). People with a higher level of conscientiousness are more likely to discuss others (Oberlander & Gill, 2006) and achievements (Hirsh & Peterson, 2009). Taken together, these findings suggest that self-reported personality traits may be associated with specific linguistic patterns in microblogs (see correlates with self-reported personality in Table 1).

2.2. Judging personality at zero-acquaintance

A large body of empirical research indicates that one's personality can be detected by unfamiliar others with reasonable accuracy. Such zero-acquaintance judgments (Kenny & West, 2008) are made possible by the presence of subtle cues, such as facial expressions (Kenny, Horner, Kashy, & Chu, 1992), physical appearance (Borkenau & Liebler, 1992; Naumann, Vazire, Rentfrow, & Gosling, 2009), choices of footwear (Gillath, Bahns, Ge, & Crandall, 2012), living environment (Gosling et al., 2002), and musical preferences (Rentfrow & Gosling, 2006).

A few studies investigated zero-acquaintance judgments in linguistic communication (see correlates with observer-judged personality in Table 1). Holleran and Mehl (2008) found that individuals could accurately judge the Big Five personality traits of unknown others by reading stream of consciousness essays. Rodriguez, Holleran, and Mehl (2010) observed accurate zero-acquaintance judgments of sub-clinical depression on the basis of self-authored descriptions. Küfner, Back, Nestler and Egloff (2010) found that raters could judge participants' openness and agreeableness via linguistic cues in creative writing samples, while Mehl et al. (2006) found that raters successfully used the presence of swear words and negative emotion words in everyday speech to judge agreeableness. Recently, with the rapid ascent of computer-mediated communication, studies have started to examine personality expression via personal websites (e.g., Vazire & Gosling, 2004), email interaction (e.g., Gill, Oberlander, & Austin, 2006), and usernames in online games (e.g., Graham & Gosling, 2012). For example, Back, Schmukle, and Egloff (2008) examined how strangers judged targets' personality through linguistic features of email addresses, and found that judgments of neuroticism, openness, agreeableness, conscientiousness, and narcissism were reasonably accurate.

2.3. The lens model of personality judgment

Brunswik's (1956) lens model provides a useful framework for conceptualizing and studying interpersonal judgment. It has been widely applied in personality judgment research (e.g., Back et al., 2010; Küfner et al., 2010; Rodriguez et al., 2010). According to the model, a given criterion variable (e.g., a personality trait such as extraversion) can be thought of as a function of several observable cues (e.g., tendency to smile, physical attractiveness). Meanwhile, the subjective judgment of that criterion variable (e.g., observer ratings of extraversion) can also be considered a function of the same cues. Cue validity is the degree of association between a given cue and the criterion variable, with a stronger correlation indicating higher validity. Cue utilization is the degree of association between a given cue and the resulting judgment, with a stronger correlation indicating greater utilization of that cue when forming personality judgments. The lens model is particularly useful because it decomposes the notion of accuracy—how closely the judgment matches the criterion variable—into two distinct parts: cue validity and cue utilization. For a personality judgment to be accurate, a cue must be (a) related to the criterion variable, and (b) successfully utilized. In essence, cues can be regarded as mediators of the criterion-judgment relationship.

The lens model offers an ideal platform for studying the relationships between microblogging, personality, and interpersonal perception. We adopt this model to examine how personality is expressed in microblogs and what linguistic cues people may use when making personality judgments.

2.4. The present research

The goals of the present study were threefold. We aimed to (1) investigate whether zero-acquaintance personality judgments can be accurately made on the basis of microbologs, (2) detect valid linguistic cues associated with personality traits, and (3) identify potential linguistic cues observers may rely onto make personality judgments. We collected all the tweets generated by an international sample of Twitter users over a 1-month period, and employed human raters and the linguistic analysis software LIWC2007 to identify personality and linguistic cues from tweets.

3. Method

3.1. Participants

We used three methods to recruit Twitter users. Firstly, we employed a snowball sampling procedure in which survey links were posted in the authors' Twitter accounts, along with requests for followers to "retweet" the survey link (i.e., disseminate the link to their own followers). We also recruited participants on-campus (in return for course credits), and through Amazon's Mechanical Turk (in return for payment of US$0.5). We stated clearly in our study description that we sought experienced Twitter users, and asked all participants to tweet a preset message so that we could verify that they were the owners of the Twitter accounts. Participants were further screened, such that only those who (a) posted more than 20 and fewer than 1000 tweets during the period from May 25th, 2011 to June 25th, 2011, and (b) posted tweets only in English, were included in the final sample.

In total, 142 participants (69 from snowball sampling, 37 from on-campus recruitment, 36 from Amazon's Mechanical Turk) were included in the following analyses. They included 55 participants from the United States (39%), 52 from Singapore (37%), 15 from the United Kingdom (11%), and 20 (13%) from other countries (i.e., Canada, India, Australia, Spain, and Oman). We also categorized the participants as either Asian ($N = 64$) and non-Asian ($N = 74$) on the basis of their self-reported ethnicity. Four participants did not indicate their ethnicity and were not classified.

3.2. Procedure

Participants filled out a two-part online survey. In the first part, participants were asked to indicate their Twitter user names, usage frequency, and to provide demographic information (i.e., gender, age, country of residence, and ethnicity). The second part comprised the 44-item Big Five Personality Inventory (BFI; John, Donahue, & Kentle, 1991). Each of the five trait measures exhibited high reliability (see Table 2).

After participants completed the survey, we verified their Twitter accounts and retrieved their tweets occurred in the past 30 days, by copying and pasting tweets directly from participants' Twitter home pages into a text file. A total of 28,978 tweets were collected between May 25th, 2011 and June 25th, 2011. On average, each participant had 204.07 tweets ($SD = 201.66$) and

Table 1
Summary of previous correlates between word categories and personality traits.

Author(s) (year)	Linguistic content	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Openness
Correlates with self-reported personality						
Back, Schmukle, and Egloff (2008)	Email address		Self-enhancing name (−)	.de, .com (−), creative name (−), cute name, funny name (−)	.net, self-enhancing name (−)	# Underscores
Golbeck, Robles, and Turner (2011)	Facebook profile	Perceptual (−), work	Affect, positive emotion, biological	Swear (−), social process, human, perceptual process (−), seeing (−)	Anxiety, ingestion	Money (−)
Gill (2003)	Email	Number (−), sport (−), affect			Inclusive, social process (−), swear (−), 1st person pronoun	
Hirsh and Peterson (2009)	Self-narrative	Human, social process, family	Certainty, inclusive, family, body (−), anger (−)	Achievement, work, body (−), death (−), anger (−), exclusive (−)	Sad, negative emotion, body, anger, home, anxiety, work (−)	Perceptual process, hear, exclusive
Holtgraves (2011)	Text message	Personal pronoun, 1st person singular pronoun, impersonal pronoun (−), anxiety (−), anger (−), sexuality, word expansion	Negative emotion (−), anxiety (−), anger (−), swear (−), death (−)		Negative emotion, acronym, emoticon	
Nguyen, Phung, Adams, and Venkatesh (2011)	Blog (Livejournal)	Word > 6 letters, leisure, number, money, perceptual, swear (−), nonfluency (−), health (−), negation (−)				
Nowson (2006)	Blog	Occupation (−), achievement (−), discrepancy (−), school (−), human, TV (−), social process	Discrepancy (−), word > 6 letters, article, negation (−)	Death (−)	Discrepancy, work, anxiety, future tense verb, eating, human (−), physical state	Word > 6 letters, positive emotion, school (−), occupation (−), grooming, inclusive, preposition
Mairesse, Walker, Mehl, and Moore (2007)	Personal essay	Article (−), body (−), certain, family, friend, 1st person singular pronoun, inclusive, music (−), negation (−), other reference, positive emotion, positive feeling, pronoun			Article (−), family, feeling, friend, human, leisure (−), music (−), negation, negative emotion, number (−), occupation (−), physical process, positive emotion (−), preposition (−), present tense verb, pronoun	
Mehl, Gosling, and Pennebaker (2006)	Daily life language usage	Word count, word > 6 letters (−)	1st person singular pronoun, article (−), swear (−)	Swear (−), negative emotion (−)	Word count (−)	Third-person pronoun (−), social process (−), past tense verb (−)
Oberlander and Gill (2006)	Email	Tentative (−), 1st person pronoun, social process, exclusive (−), inclusive, conjunction, adjective			1st person plural pronoun, 3rd person pronoun (−), inclusive, exclusive, conjunction, noun (−), adverb (−), verb, adjective	
Pennebaker and King (1999)	Personal essay	Article (−), exclusive (−), tentative (−), negation (−), inclusive, social process, positive emotion, causation (−), negative emotion (−)	1st person singular pronoun, article (−), positive emotion, negative emotion (−)	Discrepancy (−), exclusive (−), negation (−), positive emotion, causation (−), negative emotion (−)	1st person singular pronoun, article (−), positive emotion (−), negative emotion	1st person singular (−), article, word > 6 letters, present tense verb (−), exclusive, tentative, insight, causation (−)
Yarkoni (2010)	Blog	1st person plural pronoun, 2nd person pronoun, number (−), positive emotion, positive feeling, causation (−), inhibition (−), tentative (−), certainty, sensory, hearing, social process, communication, other reference, friend, family, human, up, inclusive, occupation (−), work (−), achievement (−), music, religion, physical state, sexuality	Pronoun, 1st person plural pronoun, 1st person pronoun, numbers, positive emotion, positive feeling, optimism, negative emotion (−), anger (−), causation (−), seeing, feeling, social process, other reference, friend, family, time, past tense verb, space, up, down, inclusive, motion, leisure, home, music, money (−), death (−), physical state, body state, sexuality, sleep, swear (−)	Negation (−), optimism, negative emotion (−), anger (−), sadness (−), cognitive process (−), causation (−), discrepancy (−), tentative (−), certainty (−), sensory process (−), hearing (−), human (−), time, up, exclusive (−), achievement, music (−), death (−), swear (−)	1st person singular pronoun, 2nd person pronoun (−), negation, article (−), optimism (−), negative emotion, anxiety, anger, cognitive process, causation, discrepancy, tentative, certainty, feeling, other reference (−), friend (−), space (−), up (−), exclusive, sleep, swear	Pronoun (−), 1st person singular pronoun (−), 1st person plural pronoun (−), 1st person pronoun (−), 2nd person pronoun (−), negation (−), assent (−), article, preposition, number (−), affect (−), positive emotion (−), positive feeling (−), cognitive process (−), discrepancy (−), sensory (−), social process (−), other reference (−), family (−), human (−), time (−), past tense verb (−), present tense verb (−), future tense verb

Table 1 (*continued*)

Author(s) (year)	Linguistic content	Extraversion	Agreeableness	Conscientiousness	Neuroticism	Openness
						(–), space (–), up (–), down (–), motion (–), leisure (–), home (–), sport (–), death, physical state (–), eating/drinking (–), sleep (–), grooming (–)
Yee, Harris, Jabon, and Bailenson (2011)	Text message in second life	Word > 6 letters, 1st person singular, swear (–), exclusive		Word > 6 letters, tentative	1st person singular pronoun, 2nd person pronoun, causation, discrepancy, present tense verb, future tense verb, inclusive	
Correlates with observer-judged personality						
Back, Schmukle, and Egloff (2008)	Email address	# Character (–), # dots (–), # hyphens, # digits, aol., hotmail., t-online. (–), .de (–), .com, crative name, cute name, salacious name, self-enhancing name, confusing name (–), funny name	# Characters, aol. (–), .de, cute name, self-enhancing name (–), confusing name	# Characters, # dots, # digits (–), aol. (–), t-online, .de, .net, .com (–), self-enhancing name (–), funny name (–)	# Characters, # dots, # digits (–), .de, .com (–), creative name (–), cute name, self-enhancing name (–), funny name (–)	# Dots (–), # hyphens, # digits (–), creative name, cute name
Mehl, Gosling, and Pennebaker (2006)	Daily life language usage	Sampled raw word count, swear, positive emotion, negative emotion	Word > 6 letters, 1st person singular pronoun (–), swear (–), negative emotion (–), insight, past tense verb (–)	Word > 6 letters, swear (–), nonfluency, negative emotion (–), insight	Sampled raw word count, 1st person singular pronoun (–)	Word > 6 letters, insight word, discrepancy, tentative word, social process (–), past tense verb (–)
Küfner, Back, Nestler, and Egloff (2010)	Creative writing		Positive emotion, social orientation			Sophisticated writing, creative expression, positive emotion

Note: The sign "–" represents a negative correlation between the feature and personality trait.

2362.72 words (SD = 2535.97), with a mean tweet length of 11.61 words (SD = 2.82).

After pre-processing, participants' tweets no longer contained pictures, profile information (e.g., location and gender), timestamps, or geo-locations. We further processed the tweets by removing all embedded URLs and timestamps to avoid having them contribute towards the word count. We also removed "RT" (which indicates a retweet) and words after "RT" to avoid including others' tweets in the target participants' dataset. In addition, because LIWC2007 (Pennebaker & Booth, 2007) cannot process emoticons, we replaced all the emoticons with either "PEM" (positive emoticon) or "NEM" (negative emoticon) accordingly, and added "PEM" to the positive emotion word category and "NEM" to the negative emotion word category in LIWC2007.

Then, we used LIWC2007 to generate word frequencies in our tweet samples. We split each participant's tweets into two halves by randomly selecting half of the tweets from the whole sample, and applied LIWC analysis to each half. We correlated the word frequencies and removed the categories with below moderate correlation coefficients ($r < .3$; Cohen, 1988), to ensure that the linguistic patterns that we focused on were relatively stable. This resulted in 49 out of 64 word categories being included in our final analyses.

Eight undergraduate research assistants (two males, six females; mean age = 21.75 years, SD = 2.31 years) from a large Singaporean university perused each participant's processed tweets without any time restriction, and rated their impression of the participants' personality using the same BFI that the participants used.

4. Analysis and results

4.1. Inter-observer consensus and accuracy

Inter-observer agreement was calculated using intra-class correlations for aggregate and single ratings (Vazire & Mehl, 2008). Results showed that observers had a moderate to strong agreement on extraversion, agreeableness, neuroticism, conscientiousness, and a moderate agreement on openness (see Table 2). This suggests that observers reached similar conclusions about a participant's personality, and may rely on similar cues or stereotypes when forming judgments (Graham & Gosling, 2012).

With regards to judgment accuracy, we found significant correlations between self-report and aggregated observer ratings on agreeableness and neuroticism (see Table 2). This suggests that observers can accurately predict the degree of agreeableness and neuroticism from tweets. This is consistent with the results of previous studies, which demonstrated that these two dimensions can often be judged accurately in self-related content (e.g., Holleran & Mehl, 2008). However, our results showed that observers could not accurately judge conscientiousness, extraversion, or openness.

For the accuracy of a single observer, the correlations between self-report and observer ratings were not significant and lower than the aggregate ones. This suggests a single observer cannot accurately predict the personalities of Twitter users. The use of multiple observers is needed to improve judgment accuracy.

4.2. Cue validity

In order to assess cue validity, each linguistic cue was correlated with self-reports of the Big Five personality traits. Twenty-six (13.33%) out of 245 correlations were significant at $p < .05$, exceeding chance. This suggests that the tweets contained valid linguistic cues to personality. Table 3 shows the categories that were

Table 2
Descriptive statistics for self and observer ratings of personality traits: consensus, accuracy, and vector correlations.

	Self-rating			Observer rating		Observer accuracy		Vector correlations
				Intra-class correlations (ICC)				
	M	SD	Cronbach's α	Average observer	Single observer	Aggregate observers	Single observer	
Extraversion	26.16	6.11	.83	.71***	.23***	.05	−.02	.21
Agreeableness	33.71	5.14	.76	.72***	.25***	.32**	.13	.43**
Conscientiousness	30.65	5.83	.80	.69***	.22***	.05	.02	.35*
Neuroticism	24.47	6.07	.83	.60***	.16***	.23**	.04	.74***
Openness	37.94	5.85	.80	.37***	.07***	.09	.03	.11

Note: Observer accuracy indicates the correlations between self-ratings and aggregated observers' ratings, and the average of correlations between self-ratings and a single observer's ratings. Vector correlations indicate the correlations between cue-utilization correlations and cue-validity correlations after Fisher's r-to-z transformation.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

significantly correlated with personality traits. Of the Big Five traits, only conscientiousness was found not to correlate significantly with any linguistic cues.

A number of these cue-validity correlations are consistent with previous findings. Extraversion was found to be significantly associated with social processes and positive emotion words. These results ring true given that extraverts have previously been found to use more words describing social processes in self-narratives (Hirsh & Peterson, 2009), blogs (Nowson, 2006; Yarkoni, 2010), emails (Oberlander & Gill, 2006) and stream of consciousness essays (Pennebaker & King, 1999). Extraverts have also been found to use more positive emotion words in blogs (Yarkoni, 2010) and stream of consciousness essays (Mairesse, Walker, Mehl, & Moore, 2007; Pennebaker & King, 1999). We also replicated Pennebaker and King's (1999) finding that extraversion was negatively correlated with the use of articles. These findings reflect the typical extraverts' desire for social engagement and preference for reduced complexity.

Agreeableness was found to be negatively related to negation words, an association previously observed in conventional blogging (Nowson, 2006). Neuroticism was correlated with first-person singular pronouns, replicating a previous finding in blogs (Yarkoni, 2010), and with negation words, a relationship which has previously been documented in stream of consciousness essays (Mairesse et al., 2007). Openness was negatively associated with second-person pronouns, assent words, and positive emotion words (relationships previously documented in blogs by Yarkoni (2010)), and past tense verbs (as previously observed in daily language use by Mehl et al. (2006)). This consensus reflects the consistency of personality-related behaviors across different contexts (Funder & Sneed, 1993; Gosling et al., 2002).

Our results also indicate some previously undocumented associations. For instance, we found that extraverts used more assent words, fewer functional words, and fewer impersonal pronouns. Openness was negatively related to the use of adverbs, swear words, affect words and non-fluent words, but positively related to prepositions.

4.3. Cue utilization

We correlated observers' ratings with LIWC word categories to identify possible linguistic cues individuals used when judging personality. Of 245 correlations, 70 (28.57%) were significant at $p < .05$, substantially exceeding chance (see Table 3). Ratings of extraversion were positively correlated with third-person singular pronouns, perceptual process-related words, and assent words, suggesting that observers perceived individuals who talked more about others, mentioned more of their perceptions, and expressed more agreement as more extraverted. Ratings of agreeableness were negatively correlated with swear words, negative emotion words, and anger words, suggesting that observers perceived individuals using fewer swear words and expressing less negative emotions as more agreeable. Ratings of conscientiousness were positively correlated with work-related words, and negatively with swear words and anger words, suggesting that individuals who talked more about work and less about anger were perceived as being more conscientious. The use of negative emotion words and anger words to form judgments regarding agreeableness and conscientiousness has been reported in previous studies (e.g., Küfner et al., 2010; Mehl et al., 2006).

Neuroticism ratings were correlated with swear words, negative emotion words, and anger words, suggesting that swear and negative emotion words were considered to be indicators of neuroticism by observers. Ratings of openness were positively correlated with humans, perceptual processes, and hearing word categories.

4.4. Sensitivity

The match between the pattern of cue utilization and the pattern of cue validity indicates how sensitive observers are towards using valid cues (Borkenau & Liebler, 1992). Thus, we computed vector correlations using the method proposed by Funder and Sneed (1993). First, we transformed the cue-utilization and cue-validity correlations following Fisher's r-to-z transformation. Next, we correlated the transformed correlations across all the linguistic cues for each Big Five personality dimension. The cues utilized when judging neuroticism matched closely with the valid cues to this personality trait, which may account for the observers' accuracy in judging neuroticism. Strong vector correlations also accounted for the accurate judgment of agreeableness (see Table 2). This suggests that observers used valid cues when judging neuroticism and agreeableness. Our results are consistent with previous findings in zero-acquaintance research, since accurately judged traits often exhibit high sensitivity correlations (Back et al., 2008; Borkenau & Liebler, 1992; Gosling et al., 2002).

4.5. The mediating role of linguistic cues in self-other agreement

We tested the mediating role of the linguistic cues theoretically related to the two accurately-recognized traits, agreeableness and neuroticism. We performed the mediation analysis using Preacher and Hayes' (2008) INDIRECT macro. Ninety-five percent confidence intervals and 5000 bootstrapping resamples were applied. Separate multiple mediator models were tested for agreeableness and neuroticism.

Previous studies suggest that agreeableness is positively correlated with positive emotion words and first person singular

Table 3
A Brunswik (1956) lens model analysis of judgments based on linguistic cues on Twitter: cue-validity and cue-utilization correlations.

Cue-validity correlations					Linguistic cues ("lens")	Cue-utilization correlations				
Extra.	Agree.	Cons.	Neuro.	Open.		Extra.	Agree.	Cons.	Neuro.	Open.
−.02	−.05	.01	−.09	.16	Words > 6 letters	−.17*	.06	.11	−.16	.03
_−.18_*	.02	.05	.14	−.07	Total function words	−.03	−.07	.02	_.19_*	−.02
−.08	.01	.02	.08	−.12	Total pronouns	_.18_*	−.12	−.23**	_.28_**	.01
−.02	.02	.01	.09	−.16	Personal pronouns	.26**	−.12	−.26**	_.26_**	−.01
−.11	−.04	.03	_.17_*	−.07	1st person singular (I, me, mine)	.21*	_−.23_**	−.27**	_.34_**	−.02
.06	.05	−.11	−.10	−.22**	2nd person (you, your, thou)	.08	.05	−.10	.04	.03
.11	.09	−.02	.03	−.08	3rd person plural (they, their)	_.31_**	−.05	−.15	.02	−.05
_−.19_*	−.02	.04	.02	.03	Impersonal pronouns (it, those)	−.09	−.05	−.05	.16	.06
−.18*	.04	−.06	.04	.17*	Articles (a, the)	−.19*	−.22**	−.03	.07	−.02
−.16	.06	.02	.07	−.20*	Common verbs (walk, went, see)	_.18_*	−.03	−.02	_.20_*	.04
_−.22_**	−.03	.04	.09	−.11	Auxiliary verbs (am, will, have)	.07	−.02	.06	.15	.07
−.11	−.05	.01	.15	−.20*	Past tense (went, ran, had)	.16	−.14	−.13	.13	.01
−.10	.10	.02	.00	−.14	Present tense (is, does, hear)	.22**	.00	.00	_.18_*	.05
−.11	−.03	−.04	.13	−.22**	Adverbs (very, really, quickly)	.15	.02	−.08	.10	−.04
−.05	.13	.09	−.07	.17*	Prepositions (to, with, above)	_−.30_**	−.03	.24**	−.07	−.01
−.08	−.05	.00	.17*	−.15	Conjunctions (and, but, whereas)	.08	.06	.04	.10	−.07
−.13	_−.20_*	−.10	_.20_*	−.13	Negations (no, not, never)	−.06	−.01	.03	.12	.00
−.07	.03	.07	.06	.06	Quantifiers (few, many, much)	−.09	.19*	.26**	−.11	.11
−.08	−.13	−.09	.15	−.17*	Swear words (damn, piss, fuck)	_.30_**	_−.56_**	_−.47_**	_.42_**	−.15
_.25_**	.11	.02	−.13	.06	Social processes (mate, talk, they, child)	_.26_**	_.18_*	.05	_−.20_*	.17*
−.10	−.12	−.07	−.02	.00	Humans (adult, baby, boy)	.20*	−.12	−.16	.07	.19*
_.21_*	.01	−.12	−.01	−.27**	Affective processes (happy, cried, abandon)	.22**	.13	−.19*	.11	−.02
_.28_**	.05	−.06	−.09	_−.27_**	Positive emotion (love, nice)	.22**	_.33_**	.02	−.12	.02
−.08	−.06	−.14	.16	−.07	Negative emotion (hurt, ugly, nasty)	.05	_−.33_**	_−.44_**	_.43_**	−.06
−.08	−.06	−.14	.12	−.03	Anger (hate, kill, annoyed)	.15	_−.41_**	_−.45_**	_.36_**	−.05
.05	−.05	−.08	.11	−.11	Sadness (crying, grief, sad)	.03	−.03	_−.28_**	_.22_**	−.11
−.11	.01	−.05	.10	.02	Cognitive processes (cause, know, ought)	_−.19_*	.01	_.23_**	_.17_*	.02
−.13	_−.20_*	−.01	.14	−.12	Exclusive (but, without, exclude)	−.05	−.08	.04	.13	−.11
−.05	.08	.00	.03	−.15	Perceptual processes (observing, heard, feeling)	_.31_**	−.08	_−.24_**	.06	_.28_**
.01	−.05	.11	.03	−.09	Hear (listen, hearing)	_.18_*	.03	−.15	.11	.18*
−.02	.00	−.10	.11	−.06	Biological processes (eat, blood, pain)	.12	−.13	−.31**	.17*	−.07
−.11	.00	−.09	.13	−.10	Body (cheek, hands, spit)	.08	_−.22_**	_−.32_**	.23**	−.11
−.02	_−.20_*	−.03	.15	−.01	Sexual (horny, love, incest)	_.18_*	_−.33_**	_−.34_**	.24**	.01
.06	.14	−.01	−.07	−.07	Ingestion (dish, eat, pizza)	.20*	.11	−.11	−.10	.05
−.10	.08	−.06	−.02	.12	Work (job, majors, xerox)	_−.28_**	.11	.50**	−.14	.08
−.08	−.09	.16	−.08	−.04	Leisure (cook, chat, movie)	.08	−.08	−.11	−.08	_.17_*
_.19_*	.08	−.01	−.14	−.03	Religion (altar, church, mosque)	_−.23_**	_.19_*	.12	−.09	_−.17_*
_.18_*	−.10	−.16	−.01	−.21*	Assent (agree, OK, yes)	_.31_**	.04	−.17*	−.08	.06
−.10	.01	−.08	.09	_−.21_*	Nonfluencies (er, hm, umm)	.09	.07	.03	−.03	−.12

Note: Cue-validity correlations indicate the associations between self-report ratings and word categories. Cue-utilization correlations indicate the associations between observer ratings and word categories. Only categories that correlate significantly with at least one trait are shown. Italicized values remained significant after controlling for age, gender, and ethnicity. Extra., Extraversion; Agree., Agreeableness; Cons., Conscientiousness; Neuro., Neuroticism; Open. Openness.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

pronouns, but negatively correlated with negative emotion and swear words (Mehl et al., 2006; Pennebaker & King, 1999). However, in our analysis, the indirect effect of self-reported agreeableness on observer's judgment via linguistic cue was not significant for any of the word categories.

Neuroticism was previously found to be related to greater usage of negative emotion words and less usage of positive emotion words (Mehl et al., 2006; Pennebaker & King, 1999). In the present study, the total effect of self-reported neuroticism on observers' judgment was significant ($B = .09$, $SE = .03$, $p = .005$), and the direct effect of self-reported neuroticism on observers' judgment was also significant ($B = .06$, $SE = .03$, $p = .04$). Examining the relationship between self-reported neuroticism and positive and negative emotion words revealed that self-reported neuroticism was significantly correlated with only negative emotional words ($B = .91$, $SE = .17$, $p < .001$). More importantly, of the two indirect effects, only the effect of self-reported neuroticism on observers' judgment through negative emotion was significant (CI is [.0023,.0559], excluding zero). Overall, the multiple mediator model was significant, $F(3,138) = 12.98$, $p < .001$, $R^2 = .22$. This suggests that negative emotion words partially mediated the self-other agreement correlation (i.e., accuracy) for neuroticism.

4.6. Differences in gender, age, and ethnicity

Since language style may vary with age, gender and ethnicity, we examined the possibility that the observed cue-validity correlations might be contingent on these demographic variables. We calculated the partial correlations between self-report personality and linguistic cues after controlling for age, gender, and ethnicity. The original 26 significant correlations reduced to 18 (69%), implying that some linguistic cues (e.g., past tense verbs, affective words, assent words) might not directly reflect personality, but reflect characteristics relating to age, gender, and ethnicity. Those correlations that remained significant after controlling for demographic factors likely correspond to linguistic cues that are directly related to personality traits (see italicized correlations in cue-validity in Table 3).

While our observers did not have access to participants' demographic information, participants' tweets may reveal their gender, age, or ethnicity. For example, tweets such as "I AM middle aged at 19" and "I'm totally a Bumblebee girl" explicitly indicate users' age and gender. The use of colloquial words may indicate users' nationality or ethnicity. For example, tweets such as "I like them leh haha" suggests that the user is Singaporean because the word "leh" is often used by Singaporean at the end of a sentence. Previous studies have found that observer judgment can be partially based on the stereotypes of the target's gender or age (Albright, Kenny, & Thomas, 1988; Kenny et al., 1992). To test this possibility, we followed Gillath et al.'s (2012) procedure and conducted a partial correlation between cues and observers' ratings while controlling for age, gender and ethnicity. Forty-three (61%) of the original 70 correlations remained statistically significant, which is notably fewer than the initial zero-order correlations (see italicized correlations in cue-utilization in Table 3). This suggests that observers may have relied on stereotypes of age, gender, and ethnicity when judging personality. For the two accurately judged traits, 9 (81%) out of 11 correlations for agreeableness, and 12 (80%) out of 15 correlations for neuroticism, remained statistically significant. Furthermore, the correlations between observers' ratings and self-reports of agreeableness and neuroticism remained statistically significant when controlling for age, gender, and ethnicity (agreeableness: $r = .27$, $p < .01$; neuroticism: $r = .32$, $p < .001$). This suggests that observers mainly relied on specific linguistic cues rather than stereotypes to make accurate judgments about these two traits.

5. Discussion

Previous research has documented accurate zero-acquaintance personality judgments made on the basis of writing samples, and identified the relationships between personality and language use in various contexts. The current study extends the existing findings by examining associations between microblogs and personality traits. We demonstrated that personality traits are associated with linguistic cues in microblogs and can be accurately judged by unknown others.

5.1. Personality expression

Our results showed that tweets contain valid linguistic cues to personality, and the weak to moderate correlations between linguistic cues and personality traits are comparable to those found in previous studies. In particular, extraversion was found to be positively correlated with positive emotion words and social process words, agreeableness was found to be negatively correlated with negation words, and openness was found to be negatively correlated with second-person pronouns, assent words, and positive emotion words.

Our study also identified some novel associations which may indicate the presence of other valid linguistic cues to personality. For example, extraversion was found to be negatively correlated with function words and positively correlated with assent words. Agreeableness was associated with use of fewer exclusive and sexual words. Our documentation of novel associations may be due to the fact that Twitter offers unique avenues of personality expression that are unavailable on other communication platforms. Twitter encourages individuals to talk about their daily life, share and seek information in a large network beyond a restricted group of "friends" (Java et al., 2007), and engage in frequent and prompt conversations with multiple others (Boyd, Golder, & Lotan, 2010). This combination of features cannot be found in any other computer-mediated or real-world communication platforms. It is an intuitive possibility that extraverts might communicate in a more brief and incomplete manner as they are interacting rapidly with many others, leading to omission of functional words and greater usage of assent words. Similarly, since agreeableness is associated with interpersonal concern and a desire for congruence with others, it is likely that these tendencies might be exaggerated when agreeable individuals face a large audience on Twitter, resulting in the use of fewer exclusive and potentially alienating sexual words.

We also found a number of correlations that represent more of an explanatory challenge. For example, openness was found to be negatively related to the use of adverbs, swear words, affect words and non-fluent words, but positively related to the use of prepositions. Similarly, extraversion was found to be negatively related to impersonal pronouns and auxiliary verbs. Future research is required to replicate and further investigate these findings.

5.2. Personality perception

Our results show that unfamiliar raters can accurately judge two of the Big Five dimensions, neuroticism and agreeableness, based on microblogs. This is consistent with previous findings that these two dimensions can be judged accurately in self-related content (e.g., Back et al., 2008; Holleran & Mehl, 2008). However, our results contrast with those obtained in investigations of another social media platform, Facebook. Back et al.'s (2010) study showed that observers could accurately detect extraversion, agreeableness, conscientiousness, and openness from people's Facebook profiles. This may be because Facebook profiles contain richer personal information (including self-description, status updates, and photo albums) and may provide a wider range of cues to underlying personality than tweets. However, the correlations between observer-rated and self-reported agreeableness and neuroticism were higher when forming judgments using tweets rather than Facebook profiles. This suggests that different social media platforms may afford the exhibition of different personality traits.

We also found a number of significant correlations between linguistic cues and observer-ratings, and those cues most strongly associated with the two accurately rated traits have frequently been associated with judgments of those traits in the literature (Mehl et al., 2006). Our finding that cue utilization correlations were typically stronger than cue validity correlations is also consistent with previous research (Gillath et al., 2012; Gosling et al., 2002; Mehl et al., 2006). Such asymmetry suggests that observers gave more weight to obvious characteristics that they believed to correspond with personality traits. For example, frequently talking about negative emotions is considered typical behavior for neurotic individuals.

The absence of accurate zero-acquaintance judgments of extraversion and openness may best be explained in terms of sub-optimal cue utilization. The lack of correspondence between cue validity and cue utilization correlations on the two traits suggests that individuals frequently focused on the wrong cues when judging these personality traits. The strong vector correlations for neuroticism and agreeableness add further weight to this claim, suggesting that these two traits can be judged accurately because observers exhibited more sensitivity – i.e., they employed a greater proportion of valid cues.

One possible explanation of the observed personality perception results is that microblogging affects the expression of extraversion, openness, and conscientiousness. For example, Twitter encourages people to disclose their inner feelings and share their social activities with others, meaning that all users will appear extraverted to some extent. Similarly, most people tend to tweet about their new experiences or discoveries, giving others the impression that they are open to new experiences. In fact, the ratings of openness ($M = 33.44$, $SD = 1.76$) and extraversion ($M = 28.67$, $SD = 2.54$) were the first and third highest among the ratings of the five traits, and exhibited the least variability. Judgments of conscientiousness may suffer from a lack of valid cues, since people may shy away from discussing work due to concerns over being perceived as boring. Neuroticism and agreeableness on the other hand, may be less affected by the microblogging platform. Neurotic individuals may freely disclose their negative emotions and anxiety, while less agreeable individuals may feel less need to inhibit their tendency to disagree with others. Therefore, these two traits may be easier to detect.

5.3. Theoretical and practical contributions

This research makes important theoretical and practical contributions. In terms of theoretical significance, our study is the first to examine zero-acquaintance personality judgment in the context of microblogs. It provides new empirical evidence of personality expression in social media. While much of the existing data for personality and language research has been gathered from college students in well-controlled, artificial settings, typically by laboratories in the United States (Tausczik & Pennebaker, 2009), our study extends Barker and Wright's (1951) paradigm of studying people through naturalistic observation by exploiting naturally occurring writing samples created by participants of different ages and nationalities in a natural setting using an unobtrusive approach. It presents new evidence supporting previous findings on the connection between personality and language use, while identifying new associations that have not previously been documented.

Our study also has important practical implications. With microblogging becoming increasingly popular and publicly accessible, understanding how microblogs are related to an individual's personality presents great potential for assessing personality without administrating intrusive surveys. However, it should be noted that observers' judgments showed stronger correlations with linguistic cues than self-reported personality. Hence, personality judgment based on linguistic cues from tweets is more likely to reflect the personality perceived by others, especially zero-acquaintances, instead of the true personality of the person.

A few commercial applications (e.g., Analyze Words and Tweetpsych.com) analyze tweets and produce personality reports showing how emotional and social a person is and what type of thinking style the person employs. They do not disclose their analytical methods and the accuracy of their results. While our work does not aim to provide an algorithm for automatic detection of personality from tweets, it provides empirical evidence of personality expressions in tweets, and suggests that it is possible to predict personality from tweets. However, the associations between word categories and personality are relatively weak. This suggests that software applications may need to use information beyond simple linguistic markers (e.g., social network structure) to accurately predict personality.

5.4. Limitations and future directions

One limitation of current study is that we focused on the words categorized by LIWC. This decontextualized categorization may overlook some significant higher-order semantic cues (Gill & Oberlander, 2002; Hirsh & Peterson, 2009). Future research may include more semantic analysis such as the detection of humor or writing styles to further understand the relation between personality and linguistic features.

Compared to traditional webpages or blogs, social media involves more social interaction among users. The contribution of linguistic cues alone in personality judgment might be limited. Accurate personality judgment based on social media may require not only the linguistic markers but also other behavioral cues, such as the engagement in the media and interactions with other users (Golbeck, Robles, Edmondson, & Turner, 2011; Golbeck et al., 2011; Yee, Duchenaut, Nelson, & Likarish, 2011; Yee et al., 2011). While the current study focuses on examining the relation between personality and word use, future research may include more user information in studying personality expression and judgment.

It is also possible that some word frequencies or correlations may change over time. We have attempted to minimize this possibility by using a random split-half method to exclude unstable word categories. This ensures the linguistic patterns that we analyzed are relatively stable. In addition, we distinguished the correlations that were consistent with previous findings (and are likely to be replicated) from those that had not been documented before, and attempted to provide plausible explanations for the observed cue-personality correlations. However, we do not claim that all of these correlations will be replicable. Instead, we treat them as indication of personality expression on Twitter.

While our sample size is comparable to other studies on personality and social media (e.g., Back et al., 2010), future studies should include more participants in order to verify our findings. Furthermore, we only focused on English-speaking microblogging users. As microblogging has reached millions of users in countries such as China, future research should study microblogs in other languages and identify possible cross-cultural differences in language use and personality judgment.

6. Conclusion

The present study exploits recent developments in social media to examine personality expression in naturally occurring writing samples. It replicates previous relations between personality and linguistic cues in microblogs, and identifies new associations that have not previously been documented. In addition, results show that agreeableness and neuroticism can be reliably judged by unknown others on the basis of microblog content. The study sheds light on how personality is manifested in microblogs, and offers an example of utilizing social media for personality and language research.

Acknowledgments

This research was supported by Nanyang Technological University Start-up Grant awarded to the first author. We would like to thank Cindia Toh, Natasha Laura Fong, Aloysius Seah, Wei Yang Ng, Si Hui Ong, Nina Salina, Kerk Rui Qi Yvonne and Yee Ning Lee for their help in this study.

References

Albright, L., Kenny, D. A., & Thomas, E. M. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology, 55*, 387–395.

Back, M. D., Schmukle, S. C., & Egloff, B. (2008). How extraverted is honey. bunny77@ hotmail. de? Inferring personality from e-mail addresses. *Journal of Research in Personality, 42*, 1116–1122.

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., et al. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science, 21*, 372–374.

Barker, R. G., & Wright, H. S. (1951). *One boy's day: A specimen record of behavior*. Oxford, England: Harper.

Borkenau, P., & Liebler, A. (1992). Trait inferences: Sources of validity at zero acquaintance. *Journal of Personality and Social Psychology, 62*, 645–657.

Boyd, D., Golder, S., & Lotan, G. (2010, January 5–8). *Tweet, tweet, retweet: Conversational aspects of retweeting on twitter*. Paper presented at 2010 43rd Hawaii international conference on system sciences (HICSS), Kauai, HI, USA.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.

Chiang, O. (2011, January 19). *Twitter hits nearly 200M accounts, 110M tweets per day, focuses on global expansion*. Forbes. <http://www.forbes.com/sites/oliverchiang/2011/01/19/twitter-hits-nearly-200m-users-110m-tweets-per-day-focuses-on-global-expansion/>.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Funder, D. C., & Sneed, C. D. (1993). Behavioral manifestations of personality: An ecological approach to judgmental accuracy. *Journal of Personality and Social Psychology, 64*, 479–490.

Gill, A., & Oberlander, J. (2002, August 8–10). *Taking care of the linguistic features of extraversion*. Paper presented at the 24th annual conference of the cognitive science society, Fairfax, VA.

Gill, A., & Oberlander, J. (2003, July 31-Aug 2). *Perception of email personality at zero-acquaintance. Extraversion takes care of itself; Neuroticism is a worry*. Paper presented at the Proceedings of the 25th annual conference of the cognitive science society, Boston, MA.

Gill, A., Oberlander, J., & Austin, E. (2006). Rating e-mail personality at zero acquaintance. *Personality and Individual Differences, 40*, 497–507.

Gillath, O., Bahns, A. J., Ge, F., & Crandall, C. S. (2012). Shoes as a source of first impressions. *Journal of Research in Personality, 46*, 423–430.

Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011, October 9–11). *Predicting personality from twitter*. Paper presented at the 2011 IEEE third international conference on and 2011 IEEE third international conference on social computing (SocialCom), Boston, MA, USA.

Golbeck, J., Robles, C., & Turner, K. (2011, May 7–12). *Predicting personality with social media*. Paper presented at the 2011 annual conference on human factors in computing systems, Vancouver, Canada.

Gosling, S. D., Augustine, A. A., Vazire, S., Holtzman, N., & Gaddis, S. (2011). Manifestations of personality in online social networks: Self-reported Facebook-related behaviors and observable profile information. *Cyberpsychology, Behavior, and Social Networking, 14*, 483–488.

Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology, 82*, 379–398.

Graham, L. T., & Gosling, S. D. (2012). Impressions of world of warcraft players' personalities based on their usernames: Interobserver consensus but no accuracy. *Journal of Research in Personality, 46*, 627–631.

Hirsh, J. B., & Peterson, J. B. (2009). Personality and language use in self-narratives. *Journal of Research in Personality, 43*, 524–527.

Holleran, S. E., & Mehl, M. R. (2008). Let me read your mind: Personality judgments based on a person's natural stream of thought. *Journal of Research in Personality, 42*, 747–754.

Holtgraves, T. (2011). Text messaging, personality, and the social context. *Journal of research in personality, 45*, 92–99.

Java, A., Song, X., Finin, T., & Tseng, B. (2007, August 12–15). *Why we twitter: Understanding microblogging usage and communities*. Paper presented at the 9th WebKDD and 1st SNA-KDD 2007 workshop on web mining and social network analysis, San Jose, California.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five inventory – versions 4a and 54*. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

Kenny, D. A., Horner, C., Kashy, D. A., & Chu, L. (1992). Consensus at zero acquaintance: Replication, behavioral cues, and stability. *Journal of Personality and Social Psychology, 62*, 88–97.

Kenny, D. A., & West, T. V. (2008). Self-perception as interpersonal perception. In J. V. Wood, A. Tesser, & J. G. Holmes (Eds.), *The self and social relationships* (pp. 119–137). New York: Psychology Press.

Küfner, A. C. P., Back, M. D., Nestler, S., & Egloff, B. (2010). Tell me a story and I will tell you who you are! Lens model analyses of personality and creative writing. *Journal of Research in Personality, 44*, 427–435.

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research, 30*, 457–500.

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology, 90*, 862–877.

Naaman, M., Boase, J., & Lai, C.-H. (2010, February 6–10). *Is it really about me? Message content in social awareness streams*. Paper presented at the 2010 ACM conference on computer supported cooperative work, Savannah, Georgia, USA.

Naumann, L. P., Vazire, S., Rentfrow, P. J., & Gosling, S. D. (2009). Personality judgments based on physical appearance. *Personality and Social Psychology Bulletin, 35*, 1661–1671.

Nguyen, T., Phung, D., Adams, B., & Venkatesh, S. (2011, July 17-21). *Towards discovery of influence and personality traits through social link prediction*. Paper presented at the 5th international AAAI conference on weblog and social media, Barcelona, Spain.

Nowson, S. (2006). *The Language of Weblogs: A study of genre and individual differences*. Unpublished doctoral dissertation, University of Edinburgh, Edinburgh, UK.

Oberlander, J., & Gill, A. J. (2006). Language with character: A stratified corpus comparison of individual differences in e-mail communication. *Discourse Processes, 42*, 239–270.

Oulasvirta, A., Lehtonen, E., Kurvinen, E., & Raento, M. (2010). Making the ordinary visible in microblogs. *Personal Ubiquitous Computing, 14*, 237–249.

Pennebaker, J. W., Booth, R., & Francis, M. (2007). *Linguistic inquiry and word count Mahwah*. NJ: Erlbaum.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77*, 1296–1312.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879–891.

Rentfrow, P. J., & Gosling, S. D. (2006). Message in a ballad the role of music preferences in interpersonal perception. *Psychological Science, 17*, 236–242.

Rodriguez, A. J., Holleran, S. E., & Mehl, M. R. (2010). Reading between the lines: The lay assessment of subclinical depression from written self-descriptions. *Journal of Personality, 78*, 575–598.

Rozin, P. (2001). Social psychology and science: Some lessons from Solomon Asch. *Personality and Social Psychology Review, 5*, 2–14.

Tausczik, Y. R., & Pennebaker, J. W. (2009). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*, 24–54.

Vazire, S., & Gosling, S. D. (2004). E-perceptions: Personality impressions based on personal websites. *Journal of Personality and Social Psychology, 87*, 123–132.

Vazire, S., & Mehl, M. R. (2008). Knowing me, knowing you: The accuracy and unique predictive validity of self-ratings and other-ratings of daily behavior. *Journal of Personality and Social Psychology, 95*, 1202–1216.

Yarkoni, T. (2010). Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers. *Journal of Research in Personality, 44*, 363–373.

Yee, N., Ducheneaut, N., Nelson, L., & Likarish, P. (2011, May 7-12). *Introverted elves & conscientious gnomes conscientious gnomes: the expression of personality in world of warcraft*. Paper presented at the 2011 annual conference on Human factors in computing systems, Vancouver, BC, Canada.

Yee, N., Harris, H., Jabon, M., & Bailenson, J. N. (2011). The expression of personality in virtual worlds. *Social Psychological and Personality Science, 2*, 5–12.