

# IBM Research Report

## KnowMe and ShareMe: Understanding Automatically Discovered Personality Traits from Social Media and User Sharing Preferences

**Liang Gou, Michelle X. Zhou, Huahai Yang**

IBM Almaden Research Center and IBM Global Technology Services

650 Harry Road

San Jose, CA 95120-6099

USA



Research Division

Almaden - Austin - Beijing - Cambridge - Haifa - India - T. J. Watson - Tokyo - Zurich

# KnowMe and ShareMe: Understanding Automatically Discovered Personality Traits from Social Media and User Sharing Preferences

**Liang Gou**  
IBM Research - Almaden  
San Jose, CA, USA  
lgou@us.ibm.com

**Michelle X. Zhou**  
IBM Research - Almaden  
San Jose, CA, USA  
mzhou@us.ibm.com

**Huahai Yang**  
IBM Research - Almaden  
San Jose, CA, USA  
hyang@us.ibm.com

## ABSTRACT

There is much recent work on using the digital footprints left by people on social media to predict personal traits and gain a deeper understanding of individuals. Due to the veracity of social media, imperfections in prediction algorithms, and the sensitive nature of one’s personal traits, much research is still needed to better understand the effectiveness of this line of work, including users’ preferences of sharing their computationally derived traits. In this paper, we report a two-part study involving 256 participants, which (1) examines the effectiveness of automatically deriving three types of personality traits from Twitter, including Big 5 personality, basic human values, and fundamental needs, and (2) investigates users’ opinions of using and sharing these traits. Our findings show that for over 80.8% of participants, all three types of traits derived from Twitter are significantly correlated with the participants’ psycho-metric test scores. The results also indicate over 61.5% users are willing to share their derived traits in workplace and that a number of factors significantly influence their sharing preferences. Since our findings demonstrate the feasibility and effectiveness of automatically inferring a user’s personal traits from social media, we discuss their implications for designing a new generation of privacy-preserving, hyper-personalized systems.

## Author Keywords

Personality traits, Big 5 personality, basic values, fundamental needs, privacy

## ACM Classification Keywords

H.5.2. Information Interfaces and Presentation : User Interfaces - evaluation/methodology, theory and methods, K.4.1 Computers and society: Public policy issues - privacy

## INTRODUCTION

Numerous studies from Psychology to Behavioral Economics have shown that *personality*, which is an “*individual characteristic pattern of thought, emotion, and behavior; together with the psychological mechanisms—hidden or not—behind*

*those patterns*” [11]<sup>1</sup>, influences a person’s behavior and performance in the real world, including occupational proficiency (e.g., [3]) and economic decisions (e.g., [10]). Traditionally, standard psycho-metric tests (e.g., Big 5 personality inventory) are used to gauge one’s personality, which however is impractical in many real world situations. For example, it would be difficult for a business to ask millions of its customers to take a personality test so they could receive more personalized services. Moreover, standard tests or surveys may not even exist for deriving certain personality traits, such as fundamental needs [10].

On the other hand, advances in Psycholinguistics have shown that it is feasible to *automatically* infer personality traits from one’s linguistic footprints [36, 32]. In addition, the emergence of social media has prompted hundreds of millions of people to leave their linguistic footprints on the internet. A number of research efforts have already begun to utilize these footprints to acquire a deeper understanding of individuals, including inferring their demographics and political orientation [26], Big 5 personality [12], and task fitness [20].

Inspired by these works, we are developing a system that uses one’s social media footprints (e.g., tweets) to automatically derive her personality traits. Instead of inferring personality traits based on people’s online social activities, e.g., number of posts and votes, ours analyzes the language choices in their

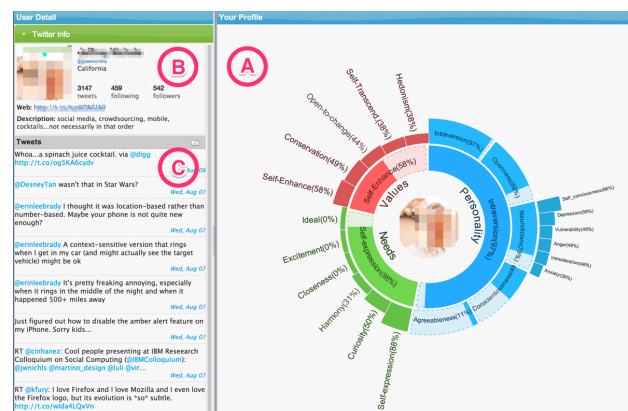


Figure 1. User interface of the experimental system, KnowMe.

<sup>1</sup>By this definition, personality traits include one’s motivation and needs. In this paper, we use “personality”, “personality traits”, and “personal traits” interchangeably to refer to three types of traits, Big 5 personality, basic values, and fundamental needs.

posts. Specifically, we develop a lexicon-based approach that predicts personality traits based on individual word choices in written samples. Our current model automatically computes three basic types of personality traits: Big 5 personality [14], fundamental needs [10], and basic values [28]. Figure 1 shows the portrait of an individual, including the three types of personality traits derived from the person’s tweets. Table 1 lists all three traits and their factors.

Due to the veracity of social media, imperfections in computational inference, and the sensitive nature of one’s personality traits, we are piloting our system with a limited group of users within our company. The main purpose is to answer two sets of questions about our system:

- How accurate are our system-derived personality traits?
  - How well do the derived traits match with the psychometric test scores?
  - How well do the derived traits match with our users’ perception about themselves?
- Whether and how would users like to use/share the derived personality traits with others in an enterprise context?
  - What and with whom users would like to share the derived traits?
  - What are the perceived benefits and risks of sharing?
  - What factors affect users’ sharing preferences of personality traits?

To answer these two sets of questions, we have designed and conducted a two-part study involving 256 participants from our company. In the first part, we derived participants’ three types of traits from their tweets and compared them with their psychometric test scores. We also solicited the participants’ perception of the derived traits. In the second part, we designed a questionnaire inquiring a participant’s preferences and concerns of sharing her derived personality traits. We then analyzed how various factors influence their preferences.

Findings from our study are promising. First, for over 80.8% of our participants, there are *statistically significant* correlations between the system-derived traits from tweets and the corresponding psychometric scores. This demonstrates the feasibility and effectiveness of automatically deriving one’s personality traits from social media. Second, over 61.5% participants are willing to share their derived traits in workplace. However, a number of factors, including the type of trait to be shared and the trait value, significantly influence their sharing preferences. This bears important implications for designing privacy-preserving, hyper-personalized systems that adapt to a user’s personality traits and privacy preferences.

## RELATED WORK

### Personality Modeling and Computation

Our work on personality modeling is grounded in Psychology and Behavioral Economics (e.g., [14, 10, 28]). However, we build computational models to *automatically* infer these traits from a person’s social media footprints.

Trait	Definition
Big5	<i>Openness-to-experience</i> (+): the extent to which a person is open to experience a variety of activities <i>Conscientiousness</i> (+): a tendency that a person acts in an organized or spontaneous way <i>Extraversion</i> (+): a tendency to seek stimulation in the company of others <i>Agreeableness</i> (+): a tendency to be compassionate and cooperative towards others <i>Neuroticism</i> (-): the extent to which a person’s emotion is sensitive to the environment
Basic Values	<i>Self-transcendence</i> (+): showing concern for the welfare and interests of others <i>Conservation</i> (-): emphasizing conformity, tradition, security <i>Self-enhancement</i> (+): seeking personal success for oneself <i>Openness-to-change</i> (+): emphasizing stimulation, self-direction <i>Hedonism</i> (-): seeking pleasure and sensuous gratification for oneself
Needs	<i>Ideals</i> (+): a desire for perfection <i>Harmony</i> (+): appreciating other people, their feelings <i>Closeness</i> (+): being connected to family and setting up home <i>Self-expression</i> (+): discovering and asserting one’s own identity <i>Excitement</i> (+): upbeat emotions, and having fun <i>Curiosity</i> (+): a desire to discover and grow

**Table 1. Three types of personality traits to share. + and - indicate a trait positively or negatively perceived in the Western Culture [19, 10, 28].**

Our approach of inferring personality traits from text builds on existing work in psycholinguistic analysis (e.g., [36, 32]). However, we extend the existing work by constructing our own psycholinguistic dictionaries to derive traits that have not been computationally modeled before, e.g., fundamental needs.

Our work is related to many efforts on inferring people’s traits from various digital footprints. They include predicting Big 5 personality from essays and conversation scripts [21] and emails [29], and inferring political orientation [26] and emotional states [7] from Twitter.

Closest to ours, Golbeck et al. use psycholinguistic analysis to predict Big 5 personality from social media (e.g., from Facebook [13] and from Twitter [12]). However, we go beyond Big 5 personality to derive two additional traits, basic values and fundamental needs, which have not been computed before. Because of our extended computational capability, our study also aims at understanding how our users would use different types of traits in their lives and the implications of such use (e.g., further loss of privacy).

### Privacy, Contextual Integrity, and Personality

There has been extensive work on understanding users’ privacy preferences of different personal data [16, 1]. Most efforts focus on data involving personal communications (e.g., instant message, email, and social media) [24, 30], and location-based activities [35, 18]. Few however have examined people’s preferences of sharing their own personality as our study does.

Another related thread of research investigates how one’s personality impacts their privacy concerns. For example, research shows that certain factors of Big 5 personality influence one’s privacy concerns [17, 19]. Compared to these works, we study how a number of factors, including Big 5, influence their privacy preferences of their personality traits.

The design of our study on understanding users’ privacy concerns is based on the contextual integrity theory [23]. It has been employed to examine information privacy (e.g., [30, 24]). However, unlike these studies, we solicit and analyze users’ opinions on sharing their own personality.

## METHOD

### Experimental System: KnowMe

We have built an experimental system, *KnowMe*, which allows a user to login with his twitter ID, collects the user’s most recent 200 public tweets, and automatically derives three types of personality traits from the tweets: Big 5 personality, basic values, and fundamental needs. Our system uses only 200 tweets, since they provide a reasonably representative sample that can produce within 10% rank of the results obtained using the person’s thousands of tweets.

We use a lexicon-based analysis based on empirically-derived correlations between personal traits and the usage of certain words or word categories. In particular, our Big 5 personality and basic values use words and word categories derived from the LIWC dictionary [36, 32], and our needs model uses a custom dictionary. To develop a custom dictionary, we used a hybrid empirical and computational approach. First, we designed and conducted large-scale, psycho-metric studies on Amazon Mechanical Turk to collect training data. For each needs dimension, we designed an item-based survey that collects a person’s psycho-metric scores describing his needs along that dimension. In the survey, we also collected participant-generated text describing a specific need. We collected the survey results from over 5000 turkers. Second, based on the collected data, we analyzed how the set of words extracted from participant-generated text are positively or negatively correlated with each needs dimension. As a result, we constructed a custom dictionary. We then built a statistical model that uses the dictionary to predict needs scores along each dimension.

Our system allows a user to interactively examine the derived personality. Figure 1 shows the system interface with three areas: three types of derived traits (panel A), twitter profile (panel B), and tweets used to derive the traits (Panel C).

### Participants

We first identified 1325 colleagues in our company who have Twitter presence and also produced at least 200 tweets. We invited each of them via email to participate in our study. 625 of them responded to our invitation, among which 256 completed the study. Among the people who completed the study, almost half of them were from the United States (42.0%), with the rest of them from Europe (32.1%) and other parts of the world (25.9%), ranging from 30 to 45 in age.

### Overall Study Design

To answer the two sets of questions posed in our Introduction, we designed and conducted a two-part study. In this study, we asked each participant to use our experimental system (*KnowMe*) and complete a two-part online survey. In the first part, our goal is to collect “ground truth” of the three types of personality traits through psycho-metric tests. The second part is to elicit users’ privacy and sharing preferences of their personality traits. On average, the whole study took about 45 minutes for a participant to complete.

### Part I. Assessing Automatically Derived Personality Traits

This study was designed to gauge the accuracy of our system-derived personality traits from Twitter, which was assessed from two aspects: comparing the derived traits with psycho-metric test scores and with users’ perception.

**Psycho-Metric Tests.** We asked each participant to take three sets of psycho-metric tests: 50-item Big 5 personality test (adopted from [www.IPIP.org](http://www.IPIP.org)), 21-item basic values test (adopted from [28]), and 52-item fundamental needs test developed by us as described above.

**Perception of Derived Traits.** Each participant was first given a video tutorial on how to use our experimental system. The participants were then asked to login with their Twitter ID and interactively explore their own personal traits derived from their tweets. Our system also provided detailed explanation for each trait. While examining the derived traits, they were also asked to rate *how well* each type of the derived trait matched with their perceptions of themselves on a five-likert scale (1 being “not at all” and 5 being “perfect”).

### Part II. Understanding Trait Sharing Preferences

Guided by the framework of contextual integrity [23], we designed a set of questions to investigate users’ privacy preferences around the four key aspects: attributes, actors, context and transmission principles.

**Attributes of information.** The *Attributes* dimension defines the type of information to be shared. In our context, we hypothesized that *trait type*, *trait value*, and *trait accuracy*, all impact users’ sharing preferences.

**Trait type.** Since the three types of personality traits reveal a person’s characteristics from different aspects, we hypothesize that people would have different privacy concerns for different traits:

- **H1a.** People have different preferences for sharing three types of personality traits.
- **H1b.** Within each type of trait, people also have different sharing preferences for its sub-traits.

**Trait value.** We also hypothesize that *trait value* would impact sharing behavior, since people may be more willing to reveal their “positive” (e.g., friendliness) but not the negative traits (e.g., emotional instability):

- **H2a.** The values of personality traits affect people’s sharing preferences.
- **H2b.** People are likely to share more high-value positive traits.

*Trait accuracy.* Previous study shows that people have shown a high privacy concern about inaccurate information of themselves [31], and have less confidence in sharing such information. We thus hypothesize:

- **H3a.** The accuracy of traits impacts people’s sharing behavior.
- **H3b.** People tend to share traits that are accurate.

*Actors.* One important factor concerning privacy is with whom the information to be shared. We hypothesize that users’ willingness to share information especially their personal traits may vary depending on the recipients of the information [4, 24]:

- **H4a.** People have different preferences about sharing their personality traits with different audience.

For each type of traits, we asked users’ sharing preferences with four groups: “public”, “distant colleagues”, “management”, and “close colleagues” [24].

Since previous work shows that certain dimensions of Big 5 personality influence one’s privacy concerns [17, 19], we also examined the participants’ own characteristics, specifically, their three types of personality traits: Big 5, basic values and needs. We hypothesize:

- **H4b.** Users’ three types of personality traits impact their sharing behavior.

Moreover, we investigated the impacts of their general disposition towards privacy and adoption of new technologies [25]. In particular, we used five questions, including three items of Disposition to Value Privacy (DTV) [34], and two items of Technology Innovativeness (INN) [2].

*Context.* Context also influences people’s privacy preferences [23]. Our context was users’ perceived benefits and risks of sharing their personality traits in an enterprise. We hypothesize that people would balance the benefits and risks when assessing privacy concerns [35]. To do so, we asked the participants to state the expected benefits and risks for sharing each type of personality trait in the work place.

*Transmission Principles.* Transmission principles indicate the types of constraints imposed on the information flow from senders to recipients. One such constraint is the granularity of information to be shared, which often impacts people’s sharing preferences [18, 9]. To test this aspect, we asked users’ preference of sharing their personality traits at three levels: “none”, “range” (ordinal scale), and “numeric” (precise score). We also sought participants’ input on desired controls for sharing their personality traits.

## RESULTS AND ANALYSES

Among 650 invited participants, 256 of them completed our study, and 224 provided complete answers to all questions. Here we report our findings to answer the two sets of questions raised in the Introduction.

### Accuracy of Personality Traits from Social Media

To answer the first set of questions, we compared the system-derived personality traits from Twitter with those from psycho-metric tests. We also gathered the participants’ perception about their derived traits.

#### *Comparing Derived Traits with Psycho-Metric Scores*

We performed correlational analyses between each type of trait and its corresponding psycho-metric scores. Within each type of trait, we performed the same analysis along each trait dimension (e.g., the five dimensions in Big 5). The results were in the range of  $0.05 < r < 0.2$ , consistent with findings from previous work [12]. Our additional analyses indicated that personality dimensions often correlate and complement each other. For example, within Big 5, *extroversion* and *agreeableness* are highly correlated ( $r=0.36, p=0.001$ ); within basic values, *conservation* and *open to change* are negatively correlated ( $r=-0.24, p=0.003$ ). We thus performed correlational analysis by considering all dimensions together within one type of trait (e.g., putting all five dimensions of Big 5 as a set of points). We then performed RV-coefficient correlation tests [27] on three sets of personality traits, respectively. All trait scores from models and questionnaires were normalized over all qualified participants.

Figure 2 shows the analysis results on all three trait models, plotting both the RV-coefficient scores and p-values against subsets of test population. The test population was ranked by the similarity between the derived scores and psychometric test scores. As the tested population increases, the RV-coefficient score decreases and p-value increases too. The vertical lines in the charts are where p-values equal to 0.05, indicating statistically significant correlations between the derived traits and their psycho-metric scores. The percentage of population at these points are 80.8%, 98.21%, and 86.6% for Big 5 personality, basic values and needs, respectively. Overall, our results indicate that for over 80% test population, our system-inferred personality traits from Twitter are significantly correlated with those from psycho-metric tests.

#### *Comparing Derived Traits with User Perception*

Each participant rated how accurate they *thought* the derived traits were. Figure 3 shows that the means of all ratings are above 3 (“somewhat”) out of 5 (“perfect”), with 3.4 ( $sd = 1.14$ ), 3.13 ( $sd = 1.17$ ), 3.39 ( $sd = 1.34$ ) for Big 5 personality, values, and needs, respectively. Although subjective, the ratings suggested that the system-derived traits somewhat matched with users’ own perception. We consider this aspect important since people often behave based on their perception (e.g., deciding on whether to share a trait depending on its accuracy).

### R2. Users’ Sharing Preferences of Personality Traits

We analyzed the effects of various factors by estimating the sharing preferences with a generalized linear mixed model with a random intercept and a log link function. The measurement is the participants’ disclosed preference (“none”, “range value” and “numeric value”) on each type of trait.

- Trait type to be shared (Big 5, basic values, and needs)
- Trait value to be shared (coded “low”, “medium”, and “high” based the ranked percentile in the test population)

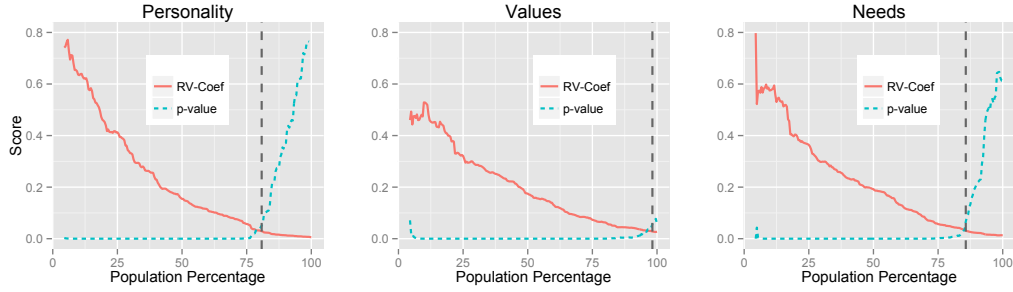


Figure 2. Accuracy of three models of personality, values and needs with RV-correlation tests.

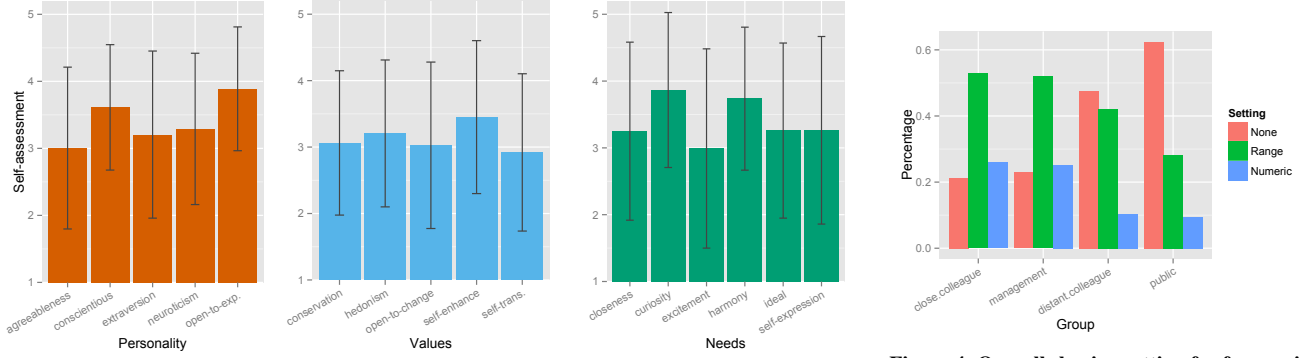


Figure 3. Average self-assessment ratings for traits of personality, values and needs.

Figure 4. Overall sharing setting for four recipient groups.

- Perceived trait accuracy (5 levels of user ratings)
- Recipient (“public”, “distant colleagues”, “management”, and “close colleagues”)
- Users’ personality traits (Big 5, basic values, and needs)
- Perceived benefits and risks (coded categories from users’ input)

Since our analysis did not identify any significant effects of the participants’ general disposition to privacy and technology innovativeness over their sharing preferences, we dropped these two factors.

*Effects of Traits*

We analyzed the effects of trait type, value, and perceived accuracy on users’ sharing preferences.

*Effects of trait type.* The type of traits impacts participants’ sharing preferences in multiple ways (Table 2). First, the type of traits has significant effect on user’s sharing preferences ( $p\text{-value}<0.001$ ). Our participants preferred to share their values trait *the least*, compared to the other two types of traits, with needs ( $\beta=-0.065, p\text{-value}<0.005$ ) and Big 5 ( $\beta=-0.08, p\text{-value}<0.001$ ). Our hypothesis **H1a** is thus supported. Such findings were corroborated by the participants’ comments on the risks of sharing one type of trait. For example, one participant stated “(values) seems VERY personal information, as it goes to the heart of what makes someone ”tick”. I think most people would feel very vulnerable if this information were shared in a work environment”. In addition, some were less willing to share traits that could not be easily observed,

Factor	Level Comparison	Hypo.
<i>Trait Type</i> ***	$PD(Values) < PD(Big5)$ *** $PD(Values) < PD(Needs)$ ***	H1a:+
<i>Big5</i> ***	$PD(Neuro.) < PD(Open.)$ *** $PD(Neuro.) < PD(Consci.)$ ** $PD(Neuro.) < PD(Extr.)$ * $PD(Neuro.) < PD(Agrb.)$ *	H1b:+
<i>Values</i>	N/A	H1b:-
<i>Needs</i>	N/A	H1b:-

Table 2. Main effects of traits on sharing preferences ( $PD(*)$  is probability of information disclosure. \*  $p<0.05$ , \*\*  $p<0.01$ , \*\*\*  $p<0.001$ )

“A few of these (needs) traits may not be easily observed in a work setting (e.g., curiosity, ideal)”.

Certain trait dimensions also impact the sharing preferences (Table 2). In Big 5, *neuroticism* has a significantly lower probability to be shared compared with other four dimensions ( $PD(*)$  indicating the expressed probability of sharing a trait). However, we did not observe any differences among trait dimensions in basic values or needs. Hypothesis **H1b** is thus supported for the Big 5 trait, but not for the other two.

*Effects of trait value.* The effects of trait value is only observed for basic values ( $p\text{-value}= 0.001$ ), but not for Big 5 nor needs. Table 3 shows the interactions among the type of trait and trait value. For Big 5, our participants preferred to disclose high *Openness*, *Conscientiousness*, and *Agreeableness*, but *not* to share high *Neuroticism*. For basic values, people preferred not to share high *Conservation* and *Hedonism*. For needs, people were willing to reveal high *Ideals*. In sum-

Type	Trait	Interaction with Score
Big 5	Openness (+)	$PD(High) > PD(Low)**$
	Conscientiousness (+)	$PD(High) > PD(Low)*$
	Agreeableness (+)	$PD(High) > PD(Low)**$
	Neuroticism (-)	$PD(High) < PD(Low)**$
Values	Conservation (-)	$PD(High) < PD(Low)*$
	Hedonism (-)	$PD(High) < PD(Low)*$
Needs	Ideal (+)	$PD(High) > PD(Low)^\circ$
	Self-Expression (+)	$PD(High) < PD(Low)^\circ$

**Table 3. Interaction among trait value and type over sharing preferences** ( $PD(*)$  is probability of information disclosure.  $^\circ p < 0.1$ ,  $* p < 0.05$ ,  $** p < 0.01$ ,  $*** p < 0.001$ )

mary, both hypotheses **H2a** and **H2b** are partially supported: trait value affects people’s sharing preferences but their preferences vary by the type of trait.

*Effects of perceived trait accuracy.* We also observed significant effects of the participants’ perceived accuracy of the derived traits on their sharing preferences across all three types of traits (all  $p$ -values  $< 0.001$ ). Our analyses showed that the participants preferred to share traits they perceived “perfect”, compared with other perceived accuracy levels (all  $\beta > 0$  and  $p$ -values  $< 0.01$ ). The preference of sharing inaccurate traits (“not at all”) is low, compared with other accuracy levels (all  $\beta < 0$  and  $p$ -values  $< 0.05$ ). This suggests that the participants were highly sensitive to the accuracy of the derived traits because inaccuracy could cause potential misunderstanding. Hypothesis **H3a** and **H3b** are thus supported.

#### Effects of Actors

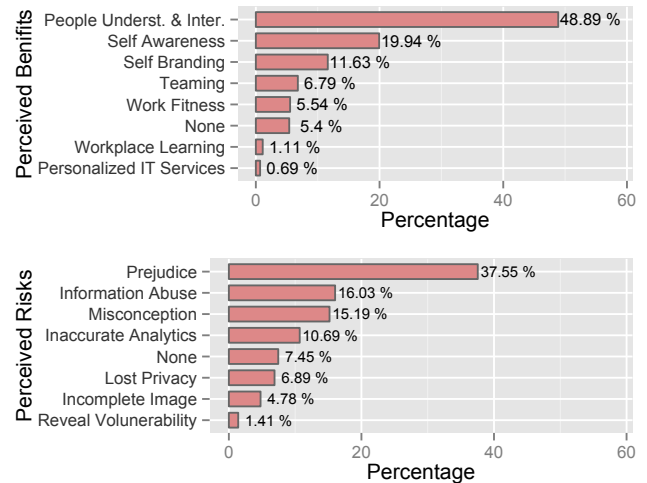
*Effects of Recipient Type.* Figure 4 shows the participants’ sharing preferences over different groups of recipients. Among all the participants, 61.5% of participants were willing to disclose their traits and share more with close colleagues and management than with others. Such sharing differences are statistically significant: ( $\beta_{\{close|management\}vs.\{distant|public\}} < 0, p < 0.001$ ). However, there was no significant difference among sharing with management and close colleagues. We did not find any interactions among the recipient factor and other factors. This suggests that the participants’ sharing preferences for different recipients were consistent across all three types of traits. Thus, hypothesis **H4a** is partially supported: recipient has main effects on people’s sharing preferences, but has no influence on the type of trait to be shared.

*Effects of Users’ Personality Traits.* We found that the participants’ personality traits<sup>1</sup> also have an impact on their sharing preferences. Table 4 shows the regression coefficients ( $\beta$ ) measuring the effects of their personality traits. As shown, certain dimensions of the participants’ personality traits significantly impact their sharing preferences. For example, *extroversion* positively impacts one’s sharing preferences for Big 5 and needs, but not for basic values. Certain dimensions consistently influences the three types of traits to be shared.

<sup>1</sup>For simplicity, we coded a participant’s personality traits derived from Twitter and his psycho-metric scores with three scales: low, medium, and high. For most of the participants, the two sets of scales are the same. If they happened to differ, we then used the one obtained from the psycho-metric tests

Factors		Types of Trait to Be Shared		
Type	Trait	Big 5	Values	Needs
Big 5	Openness-to-exp.	<b>-.075***</b>	<b>-.061**</b>	<b>-.0072***</b>
	Conscientiousness	<b>-.081***</b>	<b>-.110***</b>	<b>-.0039**</b>
	Extraversion	<b>.094***</b>	.032	<b>.0011***</b>
	Agreeableness	-.031	-.038	-.0021
	Neuroticism	-.016	-.018	-.0034 $^\circ$
Values	Hedonism	.042 $^\circ$	<b>.103***</b>	<b>.0081***</b>
	Open-to-change	.041 $^\circ$	.005	.0020
	Conservation	.041 $^\circ$	.035	-.0029
	Self-transcendence	.050 $^\circ$	.051 $^\circ$	-.0086
	Self-enhancement	<b>-.065**</b>	<b>-.086***</b>	<b>-.0058**</b>
Needs	Ideals	<b>-.116***</b>	<b>-.132***</b>	<b>-.0011***</b>
	Harmony	<b>.062*</b>	.040	.0044 $^\circ$
	Closeness	<b>.110***</b>	<b>.085***</b>	<b>.0075***</b>
	Self-expression	<b>-.062*</b>	-.051 $^\circ$	.0034
	Excitement	.038	<b>.158***</b>	.0049 $^\circ$
	Curiosity	.022	<b>.058*</b>	.0010

**Table 4. Impacts of participants’ personality traits from psychometric tests over sharing preferences** (showing  $\beta$  scores in the model.  $^\circ p < 0.1$ ,  $* p < 0.05$ ,  $** p < 0.01$ ,  $*** p < 0.001$ )



**Figure 5. Distribution of the perceived benefits and risks for sharing three types of traits.**

For example, *conscientiousness* negatively impacts the participants’ sharing preferences of all three types of traits. The hypothesis **H4b** is thus supported.

#### Perceived Benefits and Risks

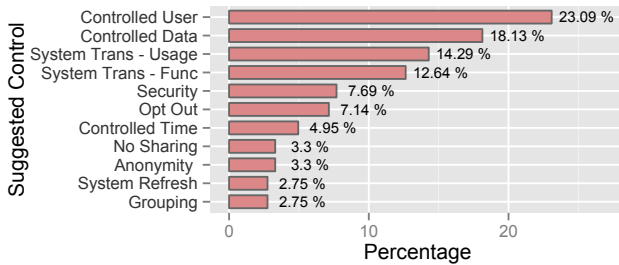
Our participants voiced a variety of benefits and risks of sharing their personality. Two coders independently read all completed responses ( $225 \times 3 \times 2 = 1344$ ) and categorized the benefits and risks. After several iterations of discussion, two coders achieved an inter-coder reliability with Cohen’s Kappa = 0.94 and 0.95 for benefits and risks, respectively.

Figure 5 shows the distribution of coded benefits and risks over three types of traits.

Table 5 shows the effects of the participants’ perceived benefits and risks on their sharing preferences. We also correlated

Type	Factor	Level Comparison of Disclosure Probability
Big 5	Benefits***	<i>Self Branding</i> > <i>Teaming</i> > { <i>Interaction, Self Awareness</i> } > <i>None</i> *
	Risks***	<i>None</i> > <i>Misconception</i> > <i>Information Abuse</i> > <i>Prejudice</i> > { <i>Inaccurate Analytics, Obsolete Information, Incomplete Image</i> } *
Values	Benefits***	{ <i>Teaming, Work Fitness</i> } > <i>Interaction</i> > <i>Self Branding</i> > <i>Self Awareness</i> > <i>None</i> *
	Risks***	<i>None</i> > <i>Information Abuse</i> > { <i>Prejudice, Misconception, Inaccurate Analytics</i> } *
Needs	Benefits***	{ <i>Interaction, Self Branding, Teaming, Work Fitness</i> } > <i>Self Awareness</i> > <i>None</i> *
	Risks***	{ <i>None, Incomplete Image, Information Abuse</i> } > { <i>Inaccurate Analytics, Lost Privacy, Misconception</i> } *

**Table 5. Effects of benefits and risks over sharing preferences (\* p<0.05, \*\* p<0.01, \*\*\* p<0.001)**



**Figure 6. Suggested controls for sharing personality traits.**

the probabilities of information disclosure with voiced benefits and risks. For example, a participant is more likely to disclose her Big 5 traits for the benefits of “*Self Branding*” than “*Teaming*”. In contrast, a participant is more likely to share her values for the benefits of “*Teaming*” and “*Work Fitness*” than “*Self Branding*”.

These effects are consistent over different trait dimensions within each type of trait, since we did not observe any interactions between benefits/risks and individual trait dimensions. However, we observed interactions among perceived benefits/risks and recipients groups. For the three types of traits, the interactions were only significant for the “*Public*” group. This suggests that the participants’ perceived benefits and risks have more impact on their sharing preferences when sharing with the “*Public*” than with other groups.

### Preferred Control Mechanisms

We also asked participants’ input about desired methods to control the sharing of personality traits. Figure 6 shows the coded results from 224 responses, including eleven categories of desired controls (Cohen’s Kappa = 0.93). As discussed below, these results provided a basis for designing a privacy-preserving, personalized system.

### IMPLICATIONS

The findings from our study present several important implications for designing a new generation of *hyper-personalized* systems, which are capable of acquiring a much deeper understanding of users based on their personal traits, and tailoring the interactions to the users based on their personality.

### Support of System Transparency

Since personality uniquely distinguishes one individual from another, many of our participants considered it very personal information. On the other hand, the participants also wished to learn about the personality of their own and others to help improve self awareness and branding, and workplace collaboration. To make such a system effective and prevent its potential abuse, our participants voiced the desire of system transparency from two main aspects: system usage and function.

First, usage wise, the system should clearly explain the meaning of each trait and intended use. For example, one participant desired “*Existence of a clear legend that states exactly what do those characteristics refer to. It might happen that people could understand something else from the name... and this should be explained very carefully*”. Users also desire concrete usage examples, “*Give some examples how others can use of misuse/misinterpret the data. Good examples with benefits and risk (as you are researching) are key*”.

Second, functionally, the system should be prescriptive and clearly states what it is capable of and its limitations. For example, one participant indicated the system’s “*ability to mark that certain attributes are inaccurate conveying the inability of system to gauge them properly*”. As another commented, “*I consider to be important to inform how many entries were taken into consideration for having such results*”, the system should also be able to explain where the results are derived from to make the results more trustworthy.

### Mixed-Initiative Privacy Preserving

Many participants (61.5%) indicated their willingness to share their personality traits with others if their privacy could be protected. Their input suggests that both users and the system should take initiatives to protect the users’ privacy.

*What to share.* Many of our participants explicitly stated that they want to control the granularity of personality traits to be shared. For example, one commented “*In this system maybe being able to switch off sections of information - maybe to allow sharing of values but not personality and needs*”. On the other hand, some participants wished the system to assess the risk of sharing certain types of information. Specifically, they want the system to provide *risk ratings* of sharing different types of traits based on “*analysis of how traits are perceived by others and the impact sharing this information might have*”. Based on the findings presented in the previous section, we believe that a system should be able to recommend proper privacy settings based on the users’ concerns (e.g., trait type and value).

*Whom to share with.* The second most voiced concern is the recipient of the information. Participants wished to specify the recipient themselves and to “*approve explicitly the list of people who will receive this information. In that way I will validate the impact of sharing the information*”.

It is interesting to note that quite a few participants wanted to be alerted or know when someone is accessing their profiles. One wished to “*see who is watching / visiting my profile*”; the other stated “*a social listening feedback loop that might be able to show a person what others might perceive or be able*”.



to derive about you from your semi-public disclosure of information. This would allow someone to be aware of what is being shared with others, and managed that as appropriate". While knowing who are accessing the information might provide a sense of protection for the information owners, it may violate the privacy of those who do not wish their system usage to be monitored and publicized. As a compromise, the system could recommend to a user with a list of potential people whom to share information with, and explain the risks and benefits of sharing information with these people.

*When to share.* Quite a few participants voiced the importance of controlling *when* to share their personality traits. For example, one participant stated "*when to share is key. We do not want to show to the new manager that I am a liberal, future thinking and creative mind if he do not appreciate that behavior. We may need to set up the background personally prior to share such results*". The ability to control when to share also provides users a sense of protection especially when they feel vulnerable: "*want to decide whether continue or stop sharing information from specific times in your life, so you can protect insights from tough personal times*". In addition to timing, participants also want to control *sharing frequency*—how often to share, an important parameter in protecting privacy. From the system standpoint, it could help a user automatically track her "downtime" and reminds the user to readjust the sharing preferences.

*Where to share.* Participants also wish to control the channel where to share their personality traits. They suggested a number of channels, from paper printout, to email, to online sharing, depending on the context. Similar to deciding on other factors, it would be helpful for the system to assess the risks of each channel and help users make informed decisions.

### **User-Assisted Personality Discovery**

In our study, it took our participants about 40 minutes on average to complete the three psychometric surveys. Our automated computational approach to deriving one's personality traits reduces users' manual efforts on taking the surveys. However, our computational models have limitations due to many factors, e.g., analytic inaccuracy, data quality, and culture influences (see Discussion section). For example, one participant stated "*My values in this field I believe is totally wrong I have always seen myself being open to change and self enhancement*". To compensate for the imperfections in analytics, we believe it is critical to allow users to interactively amend and mark the derived results. Recent research has found that such an approach is feasible and effective, since users are willing and able to correct system mistakes and improve the quality of the system over time [15].

Based on our participants' input, we have summarized several approaches for a system like ours to allow users to help improve automated personality discovery. First, the system should allow users to amend inaccurate analysis results. As our participants stated, they want the "*ability to adjust the results where you feel the system has given you an inaccurate score*". In this case, user amendments, especially collective amendments from multiple users, will help the system learn about its weaknesses and improve these areas. Second, the

system should allow users to "*comment on areas where I disagree with the results.... and let others draw a more true and pure conclusion*". Such user agreements and disagreements will help the system learn what is/is not working. Third, the system should allow users to select the data for deriving their personality traits. For example, one participant stated "*You should be able to specifically exclude specific tweets from the calculation, you might have had a tweet that was accidental or not reflective of your personality*". As indicated by the findings in [15], the "power" given to users may cause potential system abuse, where some users might manipulate the results and data for their own personal advantages (e.g., purposefully portraying a false personality). In such cases, a system must be able to detect and prevent potential misuses.

## **DISCUSSION**

### **Data Variety and Veracity**

To contain the scope of our study, we have applied psycholinguistic analysis to a person's tweets to derive his three types of personality traits: Big 5, human values, and fundamental needs. Although separate efforts have shown that Big 5 can be derived from different data sources, including essays [21], emails [29], and social media [13], it is unknown as which data sources would be most effective in deriving one's Big 5, let alone the other two types of traits in our model.

Since a person often produces different types of linguistic footprints (e.g., tweets, emails, and forum posts), a natural extension of our work is to investigate the use of different data sources in personality analysis. This is however a non-trivial task as there are many nuances in one's verbal communications depending on the context. For example, one participant in our study considered Twitter a branding channel and expresses only work-related content, "*I desire to keep my personal and professional lives separate.... and thus only share work-related content on twitter. This maintains the 'brand' image I've cultivated for years.*". On the opposite, another participant used Twitter for personal voices: "*What I say on my personal Twitter account does not necessarily reflect my personality at work*".

To make it more challenging, people's certain personality, e.g., motivations and needs, may change triggered by significant life events (e.g., becoming a parent) [10]. This implies that new data should be used to derive one's up-to-date personality. Moreover, some people maintain "multiple personalities" even using the same communication channel: "*I have been concerned though that my Work Twitter account is separate from my personal Twitter account and how would the system understand me fully? The analysis would be flawed as it would only be my work or personal account analyzed*". People may also intentionally mask their own identity, "*My twitter account is used to promote my company - I try not to put my own values or views in it*".

To gain a deeper understanding of different data sources, we see several interesting research topics. One is to collect a person's linguistic footprints from multiple sources and characterize the data by multiple dimensions (e.g., availability, veracity, and life span). We can then apply our model to col-

lected data to test the capability of different data sources. For example, we might find that emails are effective in deriving one's cognitive style while personal tweets are good for detecting emotional style. Another interesting topic is how to consolidate multiple personalities derived from different data sources. This may require a hybrid approach where both users and systems provide input to reconcile the differences in derived personalities.

### Cultural and Language Influence

Our current personality model is based on well-known psychological models developed under the Western culture. For example, our needs model is based on Maslow's hierarchy of needs, which reflects the values of twentieth-century Western middle-class males and is hardly culture neutral. In addition, our current model can only process English input. As seen from the demographics of our participants, most of them (74.1%) were rooted in the Western culture. It is thus unknown whether and how well our current work is applicable to individuals under other cultures especially a vastly different culture (e.g., Chinese).

There are numerous research efforts in investigating cultural influences on personality [33]. On the one hand, studies suggest that the basic personality traits are transcultural [22]. On the other hand, research findings already reveal much within-culture variability in Big 5 [5] and values [28]. Moreover, there exist indigeneous personality factors to a specific culture, e.g., Ren-Qing factor describing relationship orientation in Chinese culture [6].

In the meantime, the findings from our current study also suggest possible cultural influences on our results. For example, our participants' personality-sharing preferences reflected cultural influence, as they preferred to share individualist traits valued by the Western culture [33], such as openness and idealism. Moreover, some of our participants voiced how the interpretation of each trait may be influenced by culture. One participant commented "*I keep coming back to the interpretation of results. The term 'Family' for instance has very different meaning between people of different cultures and you can't apply these terms broadly and expect universal understanding or acceptance of the terms*".

To truly understand cultural influences on our model, further studies are needed. Although such studies are beyond the scope of this paper, we see several directions. One is to extend our current study to a population under a different culture (e.g., Chinese). The challenge is how we should handle the language barriers even within a similar culture. For example, one participant in our study commented "*English is not my native language and I had to rethink many times before I answered, if I had the question in my native language, there would be a possibility I would have given another answer*".

Besides cultural influences, language proficiency also influences the derived personality. As suggested by previous effort [8], we will need language-specific personality models instead of just translating a reference model (American English) into a destination language (e.g., Dutch). Second, we can also "culturize" our existing personality model by in-

corporating culture-specific personality factors (e.g., Chinese personality factors in [6]).

### CONCLUSION

In this paper, we presented a two-part study that examined (1) the accuracy of three types of personality traits derived from Twitter and (2) how users would prefer to share these traits in an enterprise setting. In the first part of study, our results show that for over 80.8% of 256 participants, all three types of traits derived from Twitter, including Big 5 personality, basic human values, and fundamental needs, are significantly correlated with the participants' corresponding psychometric test scores. In the second part of the study, our results indicate that over 61.5% users are willing to share their derived traits in workplace, and a number of factors, including the users' own personality and the perceived benefits/risks, significantly influence the users' sharing preferences. Distilled from these findings, important implications have also been discussed for guiding the design of a new generation of privacy-preserving, hyper-personalized systems.

### REFERENCES

1. Ackerman, M., Cranor, L., and Reagle, J. Privacy in e-commerce: examining user scenarios and privacy preferences. In *Proc. E-Commerce 1999* (1999), 1–8.
2. Agarwal, R., Ahuja, M., Carter, P., and Gans, M. Early and late adopters of IT innovations: extensions to innovation diffusion theory. In *Proc. DIGIT '98* (1998), 1–18.
3. Barrick, M. R., and Mount, M. K. The big five personality dimensions and job performance: a metaanalysis. *Personnel psychology* 44, 1 (1991), 1–26.
4. Bellotti, V., and Sellen, A. Design for privacy in ubiquitous computing environments. In *Proc. ECSCW '93* (1993), 77–92.
5. Bock, P. K. Culture and personality revisited. *American Behavioral Scientist* 44, 1 (2000), 32–40.
6. BOND, M. H. Localizing the imperial outreach the big five and more in chinese culture. *American Behavioral Scientist* 44, 1 (2000), 63–72.
7. De Choudhury, M., Counts, S., and Horvitz, E. Predicting postpartum changes in emotion and behavior via social media. In *Proc. CHI'2013* (2013), 3267–3276.
8. De Raad, B., Perugini, M., Hrebckov, M., and Szarota, P. Lingua franca of personality taxonomies and structures based on the psycholexical approach. *Journal of Cross-Cultural Psychology* 29, 1 (1998), 212–232.
9. Eriksson, E., Artman, H., and Swartling, A. The secret life of a persona. In *Proc. CHI' 2013* (2013), 2677–2686.
10. Ford, J. K. *Brands Laid Bare*. John Wiley & Sons, 2005.
11. Funder, D. C. *The personality puzzle*. WW Norton & Co, 1997.
12. Golbeck, J., Robles, C., Edmondson, M., and Turner, K. Predicting personality from twitter. In *Proc. SocialCom' 2011* (2011), 149–156.

13. Golbeck, J., Robles, C., and Turner, K. Predicting personality with social media. In *CHI'11 Extended Abstracts* (2011), 253–262.
14. Goldberg, L. R. The development of markers for the big-five factor structure. *Psychological Assessment* 4, 1 (1992), 26.
15. Hu, M., Yang, H., Zhou, M., Gou, L., Li, Y., and Haber, E. OpinionBlocks: a crowd-powered, self-improving interactive visual analytic system for understanding opinion text. In *Proc. INTERACT '2013* (2013), 116–134.
16. Iachello, G., and Hong, J. End-user privacy in human-computer interaction. *Foundations and Trends in Human-Computer Interaction* 1, 1 (2007), 1–137.
17. Junglas, I. A., Johnson, N. A., and Spitzmuller, C. Personality traits and concern for privacy: an empirical study in the context of location-based services. *European Journal of Information Systems* 17, 4 (2008), 387–402.
18. Knijnenburg, B., Kobsa, A., and Jin, H. Preference-based location sharing: are more privacy options really better? In *Proc. CHI '2013* (New York, NY, USA, 2013), 2667–2676.
19. Korzaan, M., Brooks, N., and Greer, T. Demystifying personality and privacy: An empirical investigation into antecedents of concerns for information privacy. *Journal of Behavioral Studies in Business* 1 (2009), 1–17.
20. Mahmud, J., Zhou, M., Megiddo, N., Nichols, J., and Drews, C. Recommending targeted strangers from whom to solicit information on social media. In *Proc. IUI '2013* (2013), 37–48.
21. Mairesse, F., Walker, M. A., Mehl, M. R., and Moore, R. K. Using linguistic cues for the automatic recognition of personality in conversation and text. *J. Artif. Intell. Res. (JAIR)* 30 (2007), 457–500.
22. McCrae, R. R., and Costa Jr, P. T. Personality trait structure as a human universal. *American psychologist* 52, 5 (1997), 509.
23. Nissenbaum, H. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press, 2009.
24. Olson, J., Grudin, J., and Horvitz, E. A study of preferences for sharing and privacy. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, ACM (2005), 1985–1988.
25. P&AB, C. P. A. A major shift since 2000 and why. *Privacy & American Business Newsletter* 10, 6 (2003).
26. Pennacchiotti, M., and Popescu, A.-M. A machine learning approach to twitter user classification. In *ICWSM* (2011).
27. Robert, P., and Escoufier, Y. A unifying tool for linear multivariate statistical methods: The RV- coefficient. *Applied Statistics* 25, 3 (1976), 257.
28. Schwartz, S. H. *Basic human values: Theory, measurement, and applications*. Revue française de sociologie, 2006.
29. Shen, J., Brdiczka, O., and Liu, J. Understanding email writers: Personality prediction from email messages. In *Proc. UMAP '2013*. Springer, 2013, 318–330.
30. Shi, P., Xu, H., and Chen, Y. Using contextual integrity to examine interpersonal information boundary on social network sites. In *Proc. CHI' 2013* (2013), 35–38.
31. Smith, H. J., and Milberg, S. J. Information privacy: measuring individuals' concerns about organizational practices. *MIS Q.* 20, 2 (June 1996), 167–196.
32. Tausczik, Y. R., and Pennebaker, J. W. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology* 29, 1 (2010), 24–54.
33. Triandis, H. C., and Suh, E. M. Cultural influences on personality. *Annual review of psychology* 53, 1 (2002), 133–160.
34. Xu, H., Dinev, T., Smith, J., and Hart, P. Information privacy concerns: linking individual perceptions with institutional privacy assurances. *Journal of the Association for Information Systems* 12, 12 (2011), 1.
35. Xu, H., Teo, H.-H., Tan, B. C., and Agarwal, R. The role of push-pull technology in privacy calculus: the case of location-based services. *Journal of Management Information Systems* 26, 3 (2009), 135–174.
36. Yarkoni, T. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality* 44, 3 (2010), 363–373.