Determining personality traits & privacy concerns from Facebook activity

Chris Sumner, Alison Byers, Matthew Shearing
The Online Privacy Foundation

Abstract

The last five years have seen a massive growth in the use of online social networks. The most popular in terms of active users is Facebook, with over 750 million members worldwide. This study explored the extent to which it is possible to determine personality traits and privacy concerns based on Facebook use. This was performed by comparing the 'Big Five' personality traits with Facebook usage, activities and language use. Results show that there are some significant correlations between an individual's personality type, their Facebook activity and their level of concern about privacy. However, the practical significance of these correlations is low. This means that making meaningful conclusions about people or taking decisions that will affect their lives on the basis of Facebook activity may therefore be problematic and error prone. These findings support and extend previous research in online social networks by showing that Facebook activity can provide limited clues to an individual's personality. However, further research into social media use is critical to ensure that the practical and ethical implications of drawing conclusions about personal information embedded in social media sites are better understood. This paper discusses online activity, personality types and privacy concerns in relation to a range of topics including marketing, pre-employment screening and susceptibility to crime such as phishing and confidence fraud.

Keywords

Personality, Social Media, Social Networks, Screening, Personnel Selection, Privacy

Acknowledgements

We would like to thank Professor Sam Gosling, University of Texas, and Gillian Empringham and Adrian Mahieu, Online Privacy Foundation, for their help and contributions to this research study.

Contact: chris@onlineprivacyfoundation.org

Introduction

The growth in online social networking has resulted in one out of every six minutes spent online being dedicated to social networking (Lipsman, 2011). Due to the massive growth in user-generated content, every two days we create as much information as we did from the dawn of civilization up until the year 2003 (Siegler, 2010). There are over 750 million active users of Facebook worldwide (Facebook, 2011) and half of the populations of the UK and USA are active Facebook users (Miniwatts Marketing Group, 2011). Despite this growth, relatively little research has been conducted into the implications of social media for society, including the potential uses of the available psychological information of its users (Gosling, Augustine, Vazire, Holtzman, & Gaddis, 2011). A lack of knowledge in this area could lead to a considerable range of ill-informed choices or misguided conclusions being made by both social media users and observers.

In the last decade we have also seen the emergence of behaviour and personality research in relation to online social networking and web log usage. A recent study of Facebook profiles (Golbeck, Robles, & Turner, 2011) clearly demonstrates a link between a user's profile information and their personality. Research into social networking use and personality is important in today's society, as knowledge of an individual's personality traits, and hence analysis of their online social networking use, could be used for a variety of purposes, including targeted marketing, employee pre-screening and fraud, including confidence crimes.

The present study aimed to examine whether online social media profiles and usage can be used as an indicator of personality traits and whether this is of practical significance. As with previous research, this study examined basic profile information such as age, sex, number of friends, biography and length of quotes. It also examined Facebook activity in greater depth, including language used in photo descriptions and wall posts. This is a key difference between this study and previous studies of personality and social media use. Language and personality have previously been examined in relation to Facebook use in the context of profile information (Golbeck, Robles, & Turner, 2011). However, it is possible that users could spend more time tailoring profile information to send signals about how they would like to be perceived by others. Wall posts and photo descriptions, however, may be more spontaneous than profile information and therefore deserve examination. This study also looked at self-reported concern over online privacy issues and whether this is related to personality type.

It is expected that some elements of Facebook activity will be significantly correlated with personality traits, as well as a person's concern over online privacy issues. It is also expected that the type of language used will correlate with personality type, but that the strength of correlation will differ between Facebook biography information, wall and photo posts, due to the spontaneity of the information provided.

Method

Participants

Five hundred and thirty-seven Facebook users from 15 countries took part in this study. Facebook profile information indicated that the majority of participants resided in Great Britain (N=291) and the United States (N=213), with 33 participants residing in other countries. The age range of participants was from 13 years to 111 years with a mean age of 30 years. Approximately two thirds of participants were female (N=349) and one third male (N=174). Fourteen participants did not report their sex. Participants volunteered for the study following advertising on Facebook, door-to-door and face-to-face leaflet distribution in Basingstoke and Cardiff (UK) and Palo Alto and Boise (USA), local press coverage in Basingstoke, and word of mouth communication. Participants were not compensated for their participation.

Materials and Design

A purpose-built Facebook application was used to collect self-reported ratings on the 44-question Big Five Inventory (John, Nauman, & Soto, 2008), providing measures of Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism. An additional question was added to capture the participants' concerns about online privacy issues. A copy of the questionnaire and how to interpret the responses can be found in **Annex A**. The application also retrieved 79 Facebook data points including sex, age, biography and quotes length and number of friends. A full list of data points can be found in **Annex B**. Some Facebook data points were restricted to a one month time period, due to restrictions in the Facebook API[1]. Specifically, at the time of writing, the Facebook API restricted users' post objects to the last 25 posts. These variables are highlighted in the annex. The application collected this data within one hour of questionnaire completion and also examined historic information. As such, participants had limited opportunity to alter their Facebook profiles and activity prior to data collection.

Data Cleansing and Processing

There were some instances in which a participant completed the questionnaire more than once. In these cases, where self-reported answers varied, all data for that participant was removed from the study. Each participant's Facebook post content was analysed using the standard categories provided in the Linguistic Inquiry and Word Count (LIWC) 2007 program.

In looking at the individual level data, it became clear that some results may be anomalous, such as very old age or very high numbers of Facebook 'friends'. Analysis was performed on the raw data, as well as on the data with outliers removed (outliers defined as more than 1.5(IQR) below Q_1 and more than 1.5(IQR) above Q_3). In comparing these two analyses, it was found that there was no difference in the significance of any results, perhaps due to the large value of N. All results in this report are based on analysis of the raw data, with no outliers removed.

[1] https://developers.facebook.com/docs/reference/api/user/

Results

The present study aimed to investigate whether there was a relationship between Facebook activity and the Big Five personality features of Extraversion (Ex), Agreeableness (Ag), Conscientiousness (Co), Neuroticism (Ne) and Openness (Op). It also looked at Facebook activity and participants' self-reported concerns over online privacy (Pr).

Privacy Concerns

A simple, zero-order Spearman's correlation was conducted on the Big Five personality features and participants' self-reported privacy concerns. These results are shown in **Table 1**.

Table 1: Spearman's correlation matrix of the Big Five personality features and self-reported privacy concerns.

	Ex	Ag	Co	Ne	Op
Ex	-				
Ag	.203 **	-			
Co	.223 **	.266 **	-		
Ne	-.289 **	-.314 **	-.318 **	-	
Op	.198 **	.090 *	-.019	-.092 *	-
Pr	-.117 **	-.095 *	-.005	.129 **	.066

** 2 tailed significance at .01 level
* 2 tailed significance at .05 level

Self-reported privacy concerns ($\mu = 3.60$, $\sigma = 1.17$) were significantly positively correlated with neuroticism ($r(535) = .129$, $p = .003$), suggesting that the more neurotic a person is, the more likely they are to be concerned with online privacy issues. Concerns over online privacy issues were also significantly negatively associated with extraversion ($r(535) = -.117$, $p = .007$) and agreeableness ($r(535) = -.111$, $p = .010$), suggesting that individuals concerned over online privacy issues are less extravert and agreeable than individuals that are less concerned with online privacy issues.

Facebook Demographics

A simple, zero-order Spearman's correlation was conducted on the Big Five personality traits, privacy concerns and Facebook demographics. These results are shown in **Table 2**.

Extraversion ($\mu = 3.30$, $\sigma = .84$) was significantly positively correlated with the number of friends a person has ($r(532) = .240$, $p < .001$) as well as the number of albums ($r(530) = .159$, $p < .001$), profile pictures ($r(521) = .146$, $p = .001$), and photos ($r(126) = .261$, $p = .003$) and post comments ($r(438) = .187$, $p < .001$) in February 2011. This suggests that the more extraverted a person is the more people they will accept as friends and will share more photographs. They are also more likely to attract comments from others on their posts. Extraversion is significantly negatively correlated with the number of books ($r(310) = -.139$, $p = .014$), suggesting that the more extraverted a person is, the less likely they are to list books on their profile.

Agreeableness ($\mu = 3.69$, $\sigma = .66$) was significantly positively correlated with age ($r(535) = .111$, $p = .010$), the number of friends a person has ($r(532) = .111$, $p = .010$) as well as the number of albums ($r(530) = .088$, $p = .042$), profile pictures ($r(521) = .093$, $p = .034$) and post comments in February 2011 ($r(438) = .097$, $p = .041$). This suggests that the older a person is, the more agreeable they are and the more friends they accept. They are also more likely to create photo albums, add profile pictures and attract comments from others on their posts.

Table 2: Spearman's correlation matrix of the Big Five, privacy concerns and Facebook demographics.

	Ex	Ag	Co	Ne	Op	Pr
Age	.053	.111 **	.229 **	-.074	.007	.066
Biography Length	.020	-.018	-.045	.079	.109 *	.005
Quotes Length	-.026	-.032	-.034	.046	.116 **	.031
Biography and Quotes Length	-.027	-.025	-.034	.075	.103 *	.020
No. Friends	.240 **	.111 *	.003	-.060	.004	-.148 **
No: Friend Lists	.006	-.040	-.066	-.046	.061	-.007
No. Activities	-.065	-.038	-.059	.089	.058	.017
No. Books	-.139 *	.050	-.023	-.045	.246 **	.005
No. Groups	.078	-.012	-.149 **	.077	.023	-.061
No. Interests	-.047	-.029	.058	-.020	.135 *	.005
No. Movies	-.053	-.053	-.040	.033	.106 *	-.022
No. Music	-.005	-.092	-.044	.088	.139 **	-.020
No. Albums	.159 **	.088 *	.059	.088 *	-.048	-.074
No. Photos	.046	-.007	.047	-.087	.020	-.011
No. Pics in Profile Pics	.146 **	.093 *	.080	.083	.008	.033
No. Photos with no Description	.090	-.062	.006	-.100	-.041	.018
No. Photos with Description	.075	.031	.064	-.099	.081	-.019
No. Photos Feb	.261 **	.150	.041	.018	.186 *	-.163
No. Posts Feb	.062	.059	-.036	.059	.092 *	.009
No. Post Comments Feb	.187 **	.097 *	.035	-.022	.035	-.093

** 2 tailed significance at .01 level
* 2 tailed significance at .05 level

Conscientiousness ($\mu = 3.39$, $\sigma = .73$) was significantly positively correlated with age ($r(535) = .229$, $p < .001$), suggesting that the older a person gets, the more conscientious they are. Conscientiousness was also significantly negatively correlated with number of groups ($r(516) = -.149$, $p = .001$), suggesting that the more conscientious an individual is, the less likely they are to join Facebook groups. However, given the correlation between conscientiousness and age, this may alternatively suggest that the number of Facebook groups an individual belongs to may be a factor of age, rather than conscientiousness alone.

Neuroticism ($\mu = 3.00$, $\sigma = .81$) was significantly positively correlated with the number of albums ($r(530) = .088$, $p = .043$), suggesting that the more neurotic a person is, the more photo albums they have on Facebook.

Openness ($\mu = 3.82$, $\sigma = .62$) was significantly positively correlated with biography length ($r(533) = .109$, $p = .012$), quotes length ($r(533) = .116$, $p = .007$), number of books ($r(310) = .246$, $p < .001$), interests ($r(229) = .135$, $p = .041$), movies ($r(387) = .106$, $p = .037$) and music ($r(449) = .139$, $p = .003$), as well as the number of photos ($r(126) = .186$, $p = .035$) and posts ($r(465) = .092$, $p = .048$) in February 2011. This suggests that the more open a person is, the more likely they are to share information about themselves, their interests and hobbies and the more likely they are to write wall posts and comment on others' posts.

Self-reported privacy concerns were significantly negatively correlated with number of fiends ($r(532) = -.148$, $p = .001$), suggesting that the more concerned with privacy an individual is, the fewer friends they will have on Facebook.

Linguistic Analysis

A simple, zero-order Spearman's correlation was conducted on the 'Big Five', privacy concerns and language used in Facebook post and profile information. These results are shown in **Table 3**. Due to the number of linguistic variables analysed, there were many significant results. This section will focus on some of the more interesting results; however, all significant results are indicated in the table with a single or double asterisk.

Extraversion was significantly positively correlated with friend words ($r(527) = .117$, $p = .007$), positive emotion words ($r(527) = .122$, $p = .005$) and assent words ($r(527) = .112$, $p = .010$), suggesting that the more extravert a person is, the more they talk about personal acquaintances. People with high extraversion are also more likely to use words indicating positive emotion, such as 'love', 'nice' or 'sweet' and indicate their assent by using words such as 'yes', 'agree' or 'OK'. Extraversion was also significantly positively correlated with words to do with biological processes ($r(527) = .095$, $p = .030$), especially words to do with ingestion ($r(527) = .139$, $p = .001$). This suggests that more extraverted people are more likely to share information about body, health and sex, but most of all information about food and eating.

Agreeableness was significantly positively associated with the number of words per sentence ($r(527) = .138$, $p = .002$), suggesting that the more agreeable a person is, the longer their sentences are, however, agreeableness is also significantly positively correlated with non-fluencies such as 'er', 'hmmm' and 'um' ($r(527) = .097$, $p = .026$), so perhaps the long sentences could be explained by these filler-type words. Agreeableness was also significantly positively correlated with positive emotion words, such as 'love', 'nice' and 'sweet' ($r(527) = .090$, $p = .039$).

Conscientiousness was significantly positively correlated with dictionary words ($r(527) = .129$, $p = .003$), suggesting that the more conscientious a person is, the more likely they are to use properly spelled words, as opposed to misspellings or text speak. Conscientiousness was also significantly positively associated with words to do with family ($r(527) = .119$, $p = .006$) and positive emotion ($r(527) = .161$, $p < .001$). Conscientiousness was negatively associated with the number of swear words used ($r(527 = -.107$, $p = .014$), suggesting that more conscientious people would swear less. It was also negatively associated with words to do with negative emotion ($r(527) = -.110$, $p < .001$), anger ($r(527) = -.139$, $p = .001$) and death ($r(527) = -.132$, $p = .002$).

Neuroticism was significantly positively correlated with the number of words used ($r(527) = .114$, $p = .009$), suggesting that the more neurotic a person is, the more they will write on social networking sites such as Facebook. Neuroticism was also positively correlated with the number of swear words used ($r(527) = .119$, $p = .006$) as well as words to do with negative emotion ($r(527) = .141$, $p = .001$), anxiety ($r(527) = .115$, $p = .008$), anger ($r(527) = .098$, $p = .024$) and sadness ($r(527) = .085$, $p = .050$), such as 'crying', 'grief' and 'sad'. It was also positively associated with biological processes ($r(527) = .162$, $p < .001$), especially words to do with the body ($r(527) = .111$, $p = .011$) and health ($r(527) = .151$, $p < .001$).

Table 3: Spearman's correlation matrix of the Big Five, privacy concerns and linguistic analysis of Facebook profiles.

	Ex	Ag	Co	Ne	Op	Pr
Word Count	.063	.010	-.017	.114 **	.151 **	.002
WPS	.016	.138 **	.058	-.009	-.087 *	.024
Words > Six Letters	-.021	-.046	-.072	-.076	.174 **	.045
Dictionary Words	.018	.054	.129 **	.039	-.049	-.013
Total Function Words	.025	.049	.101 *	.004	.024	.006
Total Pronouns	-.016	-.014	.016	-.003	.043	-.040
Personal Pronouns	.030	-.026	.009	.011	.013	-.014
1st Person Singular	.022	-.063	-.024	.023	.063	-.053
1st Person Plural	.058	.010	.085	-.037	.098 *	.042
2nd Person	-.001	.013	.043	.001	-.002	.007
3rd Person Singular	.095 *	.042	.038	.091 *	-.070	.012
3rd Person Plural	.015	.057	-.052	.047	.081	.030
Impersonal Pronouns	-.090 *	.031	.021	.008	.108 *	-.052
Articles	.076	.067	.149 **	-.068	.057	.014
Common Verbs	.017	.029	.050	.061	-.090 *	-.018
Auxiliary Verbs	.046	.067	.058	.045	-.029	-.024
Past Tense	-.026	-.049	.046	.089 *	-.015	-.037
Present Tense	.072	.065	.074	.046	-.057	-.065
Future Tense	.098 *	.003	-.002	.064	.071	.023
Adverbs	.047	.067	-.010	.078	.074	.006
Prepositions	.095 *	.036	.119 **	-.014	-.019	-.037
Conjunctions	.031	.038	.079	.063	.108 *	.023
Negations	-.047	-.040	-.068	.071	.033	-.028
Quantifiers	.017	.058	.082	.016	.065	.029
Numbers	.011	-.021	.024	.068	.130 **	.060
Swear Words	-.006	-.064	-.107 *	.119 **	.079	-.036
Social Processes	.054	.037	.073	.011	-.025	-.004
Family	.066	.042	.119 **	.013	-.128 **	-.042
Friends	.117 **	.038	.054	.025	.015	-.023
Humans	.051	.042	.049	.070	.037	.022
Affective Processes	.055	.047	.072	.082	-.043	-.048
Positive Emotion	.122 **	.090 *	.161 **	.025	-.076	-.022
Negative Emotion	-.061	-.077	-.110 *	.141 **	.089 *	-.050
Anxiety	.000	.001	.006	.115 **	.016	.002
Anger	-.068	-.058	-.139 **	.098 *	.116 **	-.068
Sadness	.018	-.006	.017	.085 *	.037	.017

** 2 tailed significance at .01 level
* 2 tailed significance at .05 level

Table 3: Spearman's correlation matrix of the Big Five, privacy concerns and linguistic analysis of Facebook profiles, continued.

	Ex	Ag	Co	Ne	Op	Pr
Cognitive Processes	-.029	.008	.027	.038	.088 *	.016
Insight	.049	-.003	-.036	.047	.053	.008
Causation	-.015	-.015	-.018	.024	.164 **	-.049
Discrepancy	.051	.032	.047	.094 *	.046	-.016
Tentative	-.067	.025	-.048	.055	.094 *	.029
Certainty	.035	.028	.040	-.001	.155 **	.027
Inhibition	-.026	.005	.035	-.002	.035	-.011
Inclusive	.081	.041	.148 **	-.016	.044	-.025
Exclusive	-.038	.008	.027	.062	.065	.020
Perceptual Processes	.031	.039	-.024	.053	.093 *	-.039
See	.085 *	.071	.021	.012	.050	.015
Hear	.013	-.026	-.019	.067	.139 **	-.026
Feel	.034	.021	.062	.112 **	.040	-.008
Biological Processes	.095 *	.022	.003	.162 **	-.056	-.056
Body	.056	.023	-.021	.111 *	-.028	-.048
Health	.015	-.025	.015	.151 **	-.008	-.020
Sexual	.064	.045	-.025	.074	.014	-.076
Ingestion	.139 **	.039	.048	.074	.000	-.035
Relativity	.087 *	.088 *	.174 **	-.013	-.059	-.025
Motion	.066	.003	.166 **	-.036	-.063	-.006
Space	.090 *	.090 *	.116 **	-.055	.035	-.025
Time	.075	.079	.120 **	.057	-.052	-.055
Work	-.039	-.050	-.053	.000	.123 **	-.055
Achievement	.076	.013	.065	-.031	.088 *	-.021
Leisure	.092 *	-.011	.029	.017	.089 *	-.067
Home	.075	.068	.102 *	.094 *	-.034	-.042
Money	.048	-.061	-.056	.029	.111 *	-.023
Religion	-.041	.038	.030	.014	.152 **	-.014
Death	-.001	-.031	-.132 **	.040	.173 **	-.026
Assent	.112 **	.020	-.053	.072	.059	-.025
Nonfluencies	.021	.097 *	-.035	.029	.079	.026
Fillers	.014	.012	-.028	.029	.081	-.004
Period	.065	-.060	.015	-.024	.166 **	.044
Comma	.000	-.006	.048	.010	.104 *	.060
Colon	.007	.046	-.024	-.026	.016	-.033
Semicolon	.089 *	.002	.045	.026	.065	.054
Question Mark	.072	.062	-.106 *	-.022	.092 *	-.078
Exclamation Mark	.166 **	.155 **	.102 *	-.016	-.082	-.046
Hyphen/Dash	-.022	-.013	-.063	.011	.070	.025
Quotation Mark	-.002	.020	.006	-.046	.195 **	-.008
Apostrophe	-.070	.056	-.102 *	.045	.077	.013
Parentheses	.030	.039	-.037	.149 **	.065	.002
Other Punctuation	-.021	-.017	-.104 *	-.012	.120 **	.001
All Punctuation	.060	.045	-.022	-.066	.126 **	.013

** 2 tailed significance at .01 level
* 2 tailed significance at .05 level

Openness was significantly positively correlated with word count ($r(527) = .151$, $p < .001$) and words greater than six letters ($r(527) = .174$, $p < .001$), suggesting that the more open a person is, the more they will write on social networking sites such as Facebook and will use longer words. However, openness was also significantly negatively correlated with the number of words per sentence ($r(527) = -.087$, $p = .047$), suggesting that even though people with higher levels of openness use more and longer words, they use shorter sentences. Openness was also significantly positively correlated with words to do with negative emotion ($r(527) = .089$, $p = .040$) and anger ($r(527) = .116$, $p = .007$), as well as traditionally taboo subjects of money ($r(527) = .111$, $p = .010$), religion ($r(527) = .152$, $p < .001$) and death ($r(527) = .173$, $p < .001$). This suggests that people with higher levels of openness, may be more open to talking about potentially sensitive subjects.

Differences in Language Use

The linguistic analysis above was based on the language used in all areas of Facebook profiles; biographies, wall posts and photo comments. The following analysis separates the language into the three separate areas to determine whether the relationships between personality types and language use differ between these three areas.

Language used in Biographies

A simple, zero-order Spearman's correlation was conducted on the Big Five personality traits, privacy concerns and language used in Facebook biographical information. These results are shown in **Annex C**.

Comparing these correlations to those of Facebook language as a whole illustrates some interesting changes.

- The relationships between extraversion and words to do with friends, positive emotion, assent and biological processes are no longer significant, suggesting that these relationships are not as applicable to biographical language as it is to other linguistic areas of Facebook.
- The relationships between agreeableness and words per sentence and words to do with positive emotion are no longer significant. There is, however, a new significant negative correlation between agreeableness and words to do with work ($r(535) = -.110$, $p = .011$), suggesting that in biographies, people with higher levels of agreeableness are less likely to discuss work, but that this relationship does not apply to all Facebook language use.
- The relationships between conscientiousness and dictionary words, swear words and words to do with family, positive emotion and anger are no longer significant. The relationship between conscientiousness and words to do with negative emotion is still significant, but less strong ($r(535) = -.085$, $p = .049$). This suggests that the biographical information of people with higher levels of conscientiousness will use more language to do with negative emotions, much like the rest of Facebook language use.
- The relationships between neuroticism and word count, swear words and words to do with negative emotion, anxiety, anger and sadness are no longer significant. This suggests that these relationships are possibly more applicable to the more spontaneous Facebook language use such as wall posts.
- The relationships between openness and words to do with family, negative emotion and anger are no longer significant. There is still a significant positive correlation between openness and word count ($r(535) = .121$, $p = .005$) and words longer than six letters ($r(535) = .109$, $p = .011$), although these relationships are not as strong. The relationship between openness and the number of words per sentence, however, has changed from a significant negative correlation to a significant positive correlation ($r(535) = .107$, $p = .013$), suggesting that people scoring higher in openness will use longer sentences in their biographical information, but shorter sentences in other areas of Facebook such as wall posts and photo descriptions. There is also a new significant positive relationship between openness and dictionary words ($r(535) = .123$, $p = .004$), suggesting that individuals with higher levels of openness will use more proper words in their biographical information, but not necessarily in other areas of Facebook.

Language used in Wall Posts

A simple, zero-order Spearman's correlation was conducted on the Big Five personality traits, privacy concerns and language used in Facebook wall posts. These results are shown in **Annex D**.

This analysis illustrated that there is no difference between the language used on Facebook wall posts and the language used across Facebook as a whole. This may be due to the volume of words used in wall posts compared to biographies or photo descriptions ($\bar{x}$ 865, 32 and 343 respectively).

Language used in Photo Descriptions

A simple, zero-order Spearman's correlation was conducted on the Big Five personality traits, privacy concerns and language used in Facebook photo descriptions. These results are shown in **Annex E**.

Similar to words used in biographies, the analysis of words used in Facebook photo posts shows some interesting differences between language used in photo descriptions and language used across Facebook as a whole.

- The correlations between extraversion and words to do with friends, positive emotion, assent and biological processes are no longer significant. There is, however, a newly significant relationship between extraversion and words to do with family ($r(318) = .157$, $p = .005$), suggesting that people with higher levels of extraversion are more likely to talk about their families when commenting on photos. This could, however, be more to do with the content of the photographs than a choice of language topic.
- The relationships between agreeableness and the number of words per sentence, nonfluencies and words to do with positive emotions are no longer significant. There are several new significant negative correlations between agreeableness and dictionary words ($r(318) = -.110$, $p = .050$), words to do with negative emotion ($r(318) = -.127$, $p = .023$), biological processes ($r(318) = -.127$, $p = .024$), achievement ($r(318) = -.112$, $p = .045$), leisure ($r(318) = -.125$, $p = .026$) and sexual words ($r(318) = -.123$, $p = .028$). This suggests that in commenting on photos, people with higher levels of agreeableness are less likely to use proper words, negative words, or words to do with biological processes, achievement, leisure activities or sex. These correlations are not significant when analysing Facebook activity as a whole, suggesting that these relationships are unique to the language used in photo comments.
- The relationships between conscientiousness and dictionary words, swear words and words to do with positive emotion, negative emotion, anger and death are no longer significant, suggesting that these relationships are more applicable to other areas of Facebook language. There is still a significant correlation between conscientiousness and words to do with family ($r(318) = -.169$, $p = .002$). This relationship is stronger in photo language than in Facebook language as a whole, suggesting that the relationship between higher levels of conscientiousness and family is more applicable to language used in photo comments. However, as suggested earlier, this may be due to the content of the photographs rather than the choice of language.
- The correlations between neuroticism and word count, swear words, and words to do with negative emotion, anger, anxiety, sadness and biological processes are no longer significant, suggesting that these relationships are more applicable to other areas of Facebook language use than to language used in photo comments.
- The correlations between openness, word count, the number of words per sentence, words longer than six letters, words to do with family, negative emotion, anger, work, money and religion are no longer significant. There is still a significant positive correlation between openness and words to do with death ($r(318) = .146$, $p = .009$), suggesting that the higher an individual's level of openness, the more likely they are to use words to do with death in photo comments.

The results from these separate correlations of the three areas of Facebook language use and their comparison with analysis of Facebook language analysis as a whole illustrate that there are some significant differences in the language used between Facebook biographies, wall posts and photo descriptions.

Discussion

Previous research has been extended in this study by looking at more data points. One of the more interesting observations was the relationship between the comments people received on their Facebook posts and their self-reported Big Five personality traits. Specifically, it was noted that people higher in scores of extraversion and agreeableness tended to attract more comments on their wall posts, with the strongest relationship being for people with higher extraversion scores. The results suggest that the more extraverted an individual is, the more comments they receive. However, as extraversion is significantly correlated with the number of friends a person has, this may suggest that the number of comments a person receives may be more to do with the number of friends they have, rather than how good they are at generating comments from others. Further analysis illustrated a significant positive relationship between the number of friends an individual has and the number of posts made in February 2011 and the number of posts comments received in February 2011 ($r(465) = .234$, $p < .000$) and $r(438) = .200$, $p < .000$ respectively), providing support to the assertion that it is the number of friends a person has that leads the number of comments they receive, rather than levels of 'extraversion'. Further analysis should be performed on this data to investigate whether it is individuals with higher scores of extraversion that lead this result, rather than simply the number of friends. Current findings do, however, support the assertion of Gosling et al (2011) that, consistent with offline behaviour, extraverts are more engaged in online social activities than introverts.

In this study, participants' self-reported online privacy concerns were also examined in relation to the Big Five personality traits and their Facebook activity. Participants with higher levels of extraversion and agreeableness showed lower levels of concern over online privacy issues. Higher levels of neuroticism, however, were associated with greater concern over online privacy issues. This was also reflected in some aspects of Facebook activity. Participants with higher levels of extraversion and agreeableness tended to have more friends listed on Facebook, whilst a concern over online privacy issues was associated with fewer Facebook friends. The association between high levels of extraversion and lower levels of privacy concerns may be explained by extraverts' tendencies to take more risks. The association between reduced privacy concerns and high agreeableness could be because people with high levels of agreeableness tend to be less suspicious and therefore more trusting that people will not misuse their personal information. The positive correlation between neuroticism and privacy concerns is unsurprising, as people higher in neuroticism are more likely to be concerned with many things, including online privacy, than those with lower levels of neuroticism. These findings raise an interesting area for future research.

Golbeck et al (Golbeck, Robles, & Turner, 2011) previously examined the linguistic usage in Facebook biographies and other profile fields and found a number of statistically significant relationships. In an extension to Golbeck's research, the current study examined the linguistic usage in wall posts and photo descriptions as well as biography information. There were several significant differences in the language use between these three Facebook areas. For example, in Facebook biographies, there are no significant relationships between any of the Big Five personality traits and words to do with family. However, in Facebook wall posts, there are, respectively, significant positive and negative relationships between words to do with family and conscientiousness and openness. In photo descriptions, on the other hand, there are significant positive relationships between words to do with family and extraversion and conscientiousness. The many differences between language use across the three areas of biographies, wall posts and photo descriptions, could be explained by users spending more time creating public biographies but giving less consideration to wall posts and photo descriptions. This could give weight to the argument that wall posts and photo descriptions are more spontaneous and therefore give stronger clues to an individual's personality traits. It could conversely be argued that biographies give the user an opportunity to reveal anything they like about themselves, whereas wall posts and photo descriptions are in reaction to a certain stimuli, therefore biographies would reveal more about an individual's personality traits than wall posts or photo descriptions. This is also a potential area for further research.

Results of this study show that there are a number of significant relationships between an individual's personality type and their Facebook activity. This presents the issue that observers of online activity could potentially reverse-

engineer a person's online activity to find clues about their personality. For example, if an individual had a relatively high number of Facebook friends, were in an older age bracket, posted a lot of information about their hobbies and interests, and wrote long sentences with language to do with friends and positive emotion, it would be reasonable to expect an observer to conclude that they were higher in extraversion and agreeableness. This type of reverse-engineering could lead to a number of conclusions being made about peoples' personalities based on their online activities. With existing research examining personality types in relation to career suitability or performance, marketing and susceptibility to fraud, reverse-engineering could prove to be a potentially dangerous area for users of social media and a potentially lucrative area for advertisers and fraudsters.

This study supports the assertion that there are a number of statistically significant relationships between personality types and Facebook activity. It also highlights that the practical relationships are relatively weak. The highest correlation found between the Big Five and Facebook demographics was the positive correlation between extraversion and the number of photos uploaded in February 2011. However, this correlation coefficient of .261 is very weak. In some contexts this may be useful to either users or observers, but in other contexts it presents an opportunity for misuse or even abuse.

The following section explores the practical implications of this study's results. Three particular areas are explored that may be of interest to social media users and society as a whole

Use Case #1 Online Marketing

Online marketing is a multi-billion dollar industry and data posted online can provide help to companies in targeting information to individuals for a number of purposes; commercial, political or social. This may be seen as being of mixed benefit for society. It has a potential benefit through improving the efficiency of commercial advertising, both for the benefit of advertisers and consumers. For example, companies such as Amazon provide recommendations based on previous purchases. However, there are risks of over-consumption and associated financial problems for individuals, especially by those with particular personality types. Indeed, such usage may be seen as manipulative and in some cases as placing unwanted constraints on the information online users receive, and thus limiting choice and access to information. Political parties are already responding to this potential, for example in campaign advertising. Moreover, information that individuals receive from search engines and Internet Service Providers (ISPs) can quite easily be filtered and limited on the basis of information that they have about users' online 'footprints'. This can be done without the users' knowledge and with or without the users' best interests at heart.

People with higher levels of agreeableness are considered to be generally less suspicious and more gullible (Costa & McCrae, 1992). This points to a greater vulnerability of this personality type both to targeted marketing and to the chances of that marketing being successful. Previous studies have also identified a link between online shopping and openness, suggesting that people higher in openness are more likely to engage in online shopping (Wang & Yang, 2008). It is also known that certain personality types prefer certain products, such as extraverts tending to like more "flashy" cars (Costa & McCrae, 2003). However, the results of the current study show that targeted marketing for people higher in agreeableness, openness and extraversion may give only some small advantage to the 'marketer', given the significant but weak correlations between online activity and personality type.

Targeted marketing of this nature may, conversely, incline observers to determine personality traits through examining the marketing that users receive. This area may warrant further research. It may also be of interest to investigate the practical significance of linking information about personality types determined from social media use to information about other forms of online usage by individuals, for example, types of purchases made, types of websites visited, etc.

Use Case #2 Online Fraud and Crimes of Confidence

A recent paper (Cisco Systems, Inc., 2011), has highlighted a rise in more targeted 'spearphishing' attacks; using contextual information to make users believe they are interacting with legitimate content. As we have already seen

in the previous use case, the Big Five personality traits of openness and agreeableness are related to an increased likelihood to engage in online marketing (higher openness) and to have lower levels of suspicion (higher agreeableness). Through what is now known about the potential of online marketing techniques it is highly likely that as the threat of cyber-attacks increases (HM Government, 2010), knowledge about personality could be used against individuals and organisations.

An interesting finding is that people with higher concern for online privacy tended to be less 'extraverted' and less 'agreeable' and tended to share less information among a smaller number of Facebook friends a lower presence on Facebook. This indicates that they could be a poor choice for a confidence trickster, who would have fewer avenues to attack (fewer friends, posts, photos etc). Their targets would also be more sceptical. This does, however, indicate a potential vulnerability for more 'extraverted' and 'agreeable' people with fewer concerns about online privacy.

Many Facebook profiles are left open and there is a tendency for more and more personal information being revealed. We are also seeing the emergence of automated tools, such as 'FBPwn' (FBPwn, 2011) to 'friend' people and then download their Facebook information. It is therefore not unreasonable to expect that fraudsters may also target certain demographics over others and with more and higher degrees of success.

In this context, our current understanding shows that there is a relationship between online social media use and personality types, but that this relationship is not strong enough on which to base meaningful conclusions alone. This presents a small advantage to criminal users at present, but future research may uncover stronger relationships which would then introduce a greater vulnerability.

Use Case #3 Employee Pre-Screening

A potential legal and civil rights minefield exists in pre-employment screening using personality derived from social media (Goldstein & Epstein, 2008). Studies of the use of personality testing in pre-employment screening have estimated that 40% of US employers (Lorenz, 2005) and 36% of British companies (Jenkins, 2001) use some form of personality testing as part of the hiring process. It's not surprising then, given the degree of personality testing in corporations (Cha, 2005) and the rise in popularity of online social networking that the use of social networks in pre-employee screening is on the rise. In 2009, 45% of employers reported that they used social media to research job candidates (Grasz, 2009) and in 2011, this figure rose to between 74% and 87% (Jobvite, 2011). To date, such screening primarily occurs in two ways; manual inspection of candidates by managers, and through using cyber-vetting companies. Screening by cyber-vetting companies is typically based on the manual analysis of profiles flagged for follow up based on certain criteria, such as keywords, affiliation with certain groups, and sexually explicit photos.

Facebook activity could also be used to screen for personality types, as recent new articles suggest (Niller, 2011). Research shows that impressions gained from online social networking profiles do influence hiring decisions (Bowie & Domke-Damonte, 2010). There is also evidence that observers are able to determine some of the personality traits of a Facebook user with reasonable accuracy (Gosling, Gaddis, & Vazire, 2007). Most troubling though, is that there is no formal research available to support the application of these approaches in hiring decisions, nor is there any consistent guidance (Brown & Vaughn, 2011).

The strength of the relationships revealed in this study indicates that there is a high probability of incorrect personality prediction from observing Facebook activity. This casts the validity of employers using such approaches as a basis for making critical employment decisions into considerable doubt.

Conclusions

This study re-affirms that there is a relationship between Facebook activity and personality types. However, it is clear that the strength of that relationship is not a strong enough basis on which to make critical decisions about individual users. The results also indicate that there may be considerable consequences in revealing personal information on Facebook. Further research is required in order to better understand the relationship between social network use and personality; the consequences; and how users might best manage the personal information they reveal through social network sites. This study points to critical questions around the possible need for regulatory controls and/or raising awareness amongst users in order to prevent the misuse of information derived from Facebook and other online social network activity.

Declaration of Conflicting Interests

The authors declare that they have no conflicts of interest with respect to their authorship and/or the publications of this article.

References

Bowie, S., & Domke-Damonte, D. (2010). Social Networking as Unconscious Impression Management: The Effect on Intention to Interview. *Southwest Academy of Management Conference.* Dallas.

Brown, V. R., & Vaughn, E. D. (2011). The Writing on the (Facebook) Wall: The Use of Social Networking Sites in Hiring Decisions. *Journal of Business and Psychology* , 219-225.

Cha, A. E. (2005, March 27). *Employers Relying on Personality Traits.* Retrieved July 4, 2011, from Washington Post: http://www.washingtonpost.com/ac2/wp-dyn/A4010-2005Mar26

Cisco Systems, Inc. (2011, June). *Email Attacks: This Time It's Personal.* Retrieved September 20, 2011, from Cisco Systems, Inc.: http://www.cisco.com/en/US/prod/collateral/vpndevc/ps10128/ps10339/ps10354/targeted_attacks.pdf

Costa, P. T., & McCrae, R. R. (2003). *Personality in adulthood: a five-factor theory perspective.* New York: The Guilford Press.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI).* Odessa, FL: Psychological Assessment Resources, Inc.

Facebook. (2011, July 4). *Facebook Statistics*. Retrieved July 4, 2011, from Facebook.com: https://www.facebook.com/press/info.php?statistics

Facebook. (2011, July 4). *Facebook Timeline*. Retrieved Juy 4, 2011, from Facebook: https://www.facebook.com/press/info.php?timeline

FBPwn. (2011, September 20). Retrieved September 20, 2011, from FBPwn - Google Code: http://code.google.com/p/fbpwn/

Golbeck, J., Robles, C., & Turner, K. (2011). Prediciting Personality with Social Media. *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems* (pp. 253-262). Vancouver: ACM.

Goldstein, A. M., & Epstein, S. D. (2008). Personality Testing in Employment: Useful Business Tool or Civil Rights Violation. *The Labor Lawyer, A Journal of Ideas and Developments in Labor and Employment Law* , 243-252.

Gosling, S. D., Gaddis, S., & Vazire, S. (2007). Personality Impressions Based on Facebook Profiles. *International Conference on Weblogs and Social Media.* Boulder, Colorado.

Gosling, S., 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* , 483-488.

Grasz, J. (2009, August 19). *Forty-five Percent of Employers Use Social Networking Sites to Research Job Candidates, CareerBuilder Survey Finds.* Retrieved September 28, 2011, from CareerBuilder: http://www.careerbuilder.com/share/aboutus/pressreleasesdetail.aspx?id=pr519&sd=8%2f19%2f2009&ed=12%2f31%2f2009&siteid=cbpr&sc_cmp1=cb_pr519_

HM Government. (2010). *Securing Britain in an Age of Uncertainty: The Strategic Defence and Security Review*. London: TSO (The Stationary Office).

Jenkins, A. (2001). *Companies' Use of Psychometric Testing and the Changing Demand for Skills: A Review of the Literature.* London: Centre for the Economics of Education.

Jobvite. (2011). *2011 Social Recruiting Survey Results.* Burlingame: Jobvite.

John, O. P., Nauman, L. P., & Soto, C. J. (2008). Paradigm Shift to the Integrative Big-Five Trait Taxonomy: History, Measurement, and Conceptual Issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), Handbook of personality: Theory and research. New York, NY: Guilford Press.

Lipsman, A. (2011, June 15). *The Network Effect: Facebook, Linkedin, Twitter & Tumblr Reach New Heights in May*. Retrieved September 20, 2011, from comScore: http://blog.comscore.com/2011/06/facebook_linkedin_twitter_tumblr.html

Lorenz, K. (2005, February 25). *Personality tests help gauge job fit*. Retrieved September 28, 2011, from CNN: http://edition.cnn.com/2005/US/Careers/02/25/personality.tests/

Miniwatts Marketing Group. (2011, June 30). *Facebook Usage and Facebook Penetration Statistics By World Geographic Regions*. Retrieved September 20, 2011, from Internet World Stats: http://www.internetworldstats.com/facebook.htm

Contact: chris@onlineprivacyfoundation.org

Nicholson, N., Soane, E., Fenton-O'Creevy, M., & Willman, P. (2005). Personality and domain-specific risk-taking. *Journal of Risk Research* , 157-176.

Niller, E. (2011, May 9). *Facebook Can Serve as Personality Test*. Retrieved September 28, 2011, from Discovery News: http://news.discovery.com/tech/facebook-personality-test-employers-110509.html

Siegler, M. (2010, August 4). *Eric Schmidt: Every 2 Days We Create As Much Information As We Did Up To 2003*. Retrieved September 20, 2011, from Tech Crunch: http://techcrunch.com/2010/08/04/schmidt-data/

Social Intelligence Corp. (2011, September 20). *Social Intelligence*. Retrieved September 20, 2011, from Social Intelligence: http://www.socialintel.com/home

Wang, C.-C., & Yang, H.-W. (2008). PASSION FOR ONLINE SHOPPING: THE INFLUENCE OF PERSONALITY AND COMPULSIVE BUYING. *Social Behavior and Personality: an international journal* , 693-706.

Annex A

The Big Five Inventory

How I am in general

Here are a number of characteristics that may or may not apply to you. For example, do you agree that you are someone who *likes to spend time with others*? Please write a number next to each statement to indicate the extent to which **you agree or disagree with that statement.**

1	**2**	**3**	**4**	**5**
Disagree Strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly

I am someone who…

1. _____ Is talkative
2. _____ Tends to find fault with others
3. _____ Does a thorough job
4. _____ Is depressed, blue
5. _____ Is original, comes up with new ideas
6. _____ Is reserved
7. _____ Is helpful and unselfish with others
8. _____ Can be somewhat careless
9. _____ Is relaxed, handles stress well.
10. _____ Is curious about many different things
11. _____ Is full of energy
12. _____ Starts quarrels with others
13. _____ Is a reliable worker
14. _____ Can be tense
15. _____ Is ingenious, a deep thinker
16. _____ Generates a lot of enthusiasm
17. _____ Has a forgiving nature
18. _____ Tends to be disorganized
19. _____ Worries a lot
20. _____ Has an active imagination
21. _____ Tends to be quiet
22. _____ Is generally trusting
23. _____ Tends to be lazy

24. _____ Is emotionally stable, not easily upset

25. _____ Is inventive

26. _____ Has an assertive personality

27. _____ Can be cold and aloof

28. _____ Perseveres until the task is finished

29. _____ Can be moody

30. _____ Values artistic, aesthetic experiences

31. _____ Is sometimes shy, inhibited

32. _____ Is considerate and kind to almost everyone

33. _____ Does things efficiently

34. _____ Remains calm in tense situations

35. _____ Prefers work that is routine

36. _____ Is outgoing, sociable

37. _____ Is sometimes rude to others

38. _____ Makes plans and follows through with them

39. _____ Gets nervous easily

40. _____ Likes to reflect, play with ideas

41. _____ Has few artistic interests

42. _____ Likes to cooperate with others

43. _____ Is easily distracted

44. _____ Is sophisticated in art, music, or literature

Privacy Question

I am someone who…

45. _____ Is concerned about privacy issues

Scoring Instructions

To score the BFI, you'll first need to **reverse-score** all negatively-keyed items:

Extraversion:	6, 21, 31
Agreeableness:	2, 12, 27, 37
Conscientiousness:	8, 18, 23, 43
Neuroticism:	9, 24, 34
Openness:	35, 41

To recode these items, you should subtract your score for all reverse-scored items from 6. For example, if you gave yourself a 5, compute 6 minus 5 and your recoded score is 1. That is, a score of 1 becomes 5, 2 becomes 4, 3 remains 3, 4 becomes 2, and 5 becomes 1.

Next, you will create scale scores by averaging the following items for each B5 domain (where R indicates using the reverse-scored item).

Extraversion:	1, 6R 11, 16, 21R, 26, 31R, 36
Agreeableness:	2R, 7, 12R, 17, 22, 27R, 32, 37R, 42
Conscientiousness:	3, 8R, 13, 18R, 23R, 28, 33, 38, 43R
Neuroticism:	4, 9R, 14, 19, 24R, 29, 34R, 39
Openness:	5, 10, 15, 20, 25, 30, 35R, 40, 41R, 44

SPSS Syntax

```
*** REVERSED ITEMS

RECODE
 bfi2 bfi6 bfi8 bfi9 bfi12 bfi18 bfi21 bfi23 bfi24 bfi27 bfi31 bfi34 bfi35
 bfi37 bfi41 bfi43
 (1=5) (2=4) (3=3) (4=2) (5=1) INTO bfi2r bfi6r bfi8r bfi9r bfi12r bfi18r bfi21r bfi23r bfi24r
 bfi27r bfi31r bfi34r bfi35r bfi37r bfi41r bfi43r.
EXECUTE .

*** SCALE SCORES

COMPUTE bfie = mean(bfi1,bfi6r,bfi11,bfi16,bfi21r,bfi26,bfi31r,bfi36) .
VARIABLE LABELS bfie 'BFI Extraversion scale score.
EXECUTE .

COMPUTE bfia = mean(bfi2r,bfi7,bfi12r,bfi17,bfi22,bfi27r,bfi32,bfi37r,bfi42) .
VARIABLE LABELS bfia 'BFI Agreeableness scale score' .
EXECUTE .

COMPUTE bfic = mean(bfi3,bfi8r,bfi13,bfi18r,bfi23r,bfi28,bfi33,bfi38,bfi43r) .
VARIABLE LABELS bfic 'BFI Conscientiousness scale score' .
EXECUTE .

COMPUTE bfin = mean(bfi4,bfi9r,bfi14,bfi19,bfi24r,bfi29,bfi34r,bfi39) .
VARIABLE LABELS bfin 'BFI Neuroticism scale score' .
EXECUTE .

COMPUTE bfio = mean(bfi5,bfi10,bfi15,bfi20,bfi25,bfi30,bfi35r,bfi40,bfi41r,bfi44) .
VARIABLE LABELS bfio 'BFI Openness scale score' .
EXECUTE .
```

Reference Information

The BFI should be cited with the original and a more accessible, recent reference:

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.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), Handbook of personality: Theory and research (pp. 114-158). New York, NY: Guilford Press.

Annex B

Table B1: Data points collected by the Big Five Facebook application.

No:	Facebook Object Fields	Data Point Collected	Data Points used in Analysis?	Notes
1	User ID	User ID	No	
2	Name	Name	No	
3		Length of Name	No	
4	First Name	First Name	No	
5		Length of First Name	Yes	
6	Last Name	Last Name	No	
7		Length of Last Name	Yes	
8	URL	URL	Yes	
9	Vanity URL	Vanity URL Yes/No	Yes	
10	Date of Birth	Date of birth	No	
11		Age	Yes	
12	Gender	Gender	Yes	
13	EmailAddress	Email Address	No	
14		Length of Email Address	Yes	
15	Locale	Locale	Yes	
16	Verified	Verified	No	
17	Updated	Updated	No	
18	Biography	Biography text	Yes	In LIWC
19		Biography Length (Number of characters)	Yes	
20	Quotes	Quotes Text	Yes	In LIWC
21		Quotes Length (Number of characters)	Yes	
22		Combined Biography and Quotes Length	Yes	
23		Relationship Status	Yes	
24		Political Views		
25	Friends	Friend ID	No	
26		Friend Name	No	
27		Number of friends	Yes	
28	Friend Lists	Friend List ID	No	
29		FriendList Name	No	
30		Number of FriendsLists	Yes	
31	Activity	Activity ID	No	
32		Activity Name	No	
33		Number of Activities	Yes	
34	Books	Book ID	No	
35		Book Name	No	
36		Number of Books	Yes	
37	Groups	Group ID	No	

Table B1: Data points collected by the Big Five Facebook application continued.

No:	Facebook Object Fields	Data Point Collected	Data Points used in Analysis?	Notes
38		Group Name	No	
39		Number of Groups	Yes	
40	Interests	Interest ID	No	
41		Interest Name	No	
42		Number of Interests	Yes	
43	Movies	Movie ID	No	
44		Movie Name	No	
45		Number of Movies	Yes	
46	Music	Music ID	No	
47		Music Name	No	
48		Number of Music	Yes	
49	Television	Television ID	No	
50		Television Name	No	
51		Number of Television	Yes	
52	Photo Albums	Album ID	No	
53		Album Name	Yes	
54		Album Privacy	No	
55		Album Create Date	No	
56		Number of Photo Albums	Yes	
57	Photos	Album ID	No	
58		Photo URL	No	
59		Photo Description	Yes	
60		Photo Created Date/Time	No	
61		Photo Comments	No	
62		Number of Photos	Yes	
63		Number of Photos in Profile Pictures Album	Yes	
64		Number of Photos with a Description	Yes	
65		Number of Photos without descriptions	Yes	
66		Number of Photos uploaded during February 2011	Yes	Limited to 1 month
67		LIWC analysis of Photo Descriptions	Yes	
68		Analysis of the use of LOL in photo descriptions during February 2011	No	Limited to 1 month

Table B1: Data points collected by the Big Five Facebook application continued.

No:	Facebook Object Fields	Data Point Collected	Data Points used in Analysis?	Notes
69	Posts	Post ID	No	
70		Post message	Yes	
71		Post Created Date/Time	No	
72		Post Comments	Yes	
73		Post likes	No	
74		Number of Posts	Yes	
75		Number of Posts during February 2011	Yes	Limited to 1 month
76		Number of comments on Posts during February 2011	Yes	Limited to 1 month
77		Word count from posts in February 2011	Yes	Limited to 1 month
78		LIWC analysis of post messages during February 2011	Yes	Limited to 1 month
79		Analysis of the use of LOL in posts during February 2011	No	Limited to 1 month

Annex C

Table C1: Spearman's correlation matrix of the Big Five, privacy concerns and linguistic analysis of Facebook biographies.

	Ex	Ag	Co	Ne	Op	Pr
Word Count	.014	.004	-.045	.059	.121 **	-.003
WPS	.013	.018	-.052	.046	.107 *	-.012
Words > Six Letters	.022	-.029	-.045	.077	.109 *	.016
Dictionary Words	.062	.020	-.063	.073	.123 **	-.024
Total Function Words	.031	-.007	-.051	.074	.129 **	-.025
Total Pronouns	.053	.028	-.080	.087 *	.108 *	-.023
Personal Pronouns	.069	.046	-.072	.091 *	.118 **	-.020
1st Person Singular	.070	.058	-.038	.048	.122 **	-.004
1st Person Plural	-.044	-.058	-.028	.037	-.026	.018
2nd Person	.008	.026	-.081	.074	.031	-.053
3rd Person Singular	-.024	.031	.009	.062	.086 *	-.003
3rd Person Plural	-.019	-.014	.084	.009	.050	-.012
Impersonal Pronouns	.005	-.062	-.025	.071	.038	-.014
Articles	-.011	-.029	-.009	.055	.129 **	.021
Common Verbs	.021	.012	-.044	.073	.094 *	-.062
Auxiliary Verbs	.003	-.005	-.017	.061	.076	-.047
Past Tense	-.023	-.044	.006	.062	.083	.013
Present Tense	.037	.008	.000	.081	.081	-.061
Future Tense	.009	-.084	-.037	.080	.059	-.054
Adverbs	.019	-.024	-.037	.060	.100 *	-.033
Prepositions	.005	-.026	-.065	.066	.087 *	-.035
Conjunctions	.019	-.080	-.027	.064	.125 **	-.029
Negations	-.016	-.030	-.041	.123 **	.115 **	-.055
Quantifiers	-.002	-.032	.001	.104 *	.089 *	.026
Numbers	-.092 *	.000	-.007	.081	.005	.029
Swear Words	.003	-.052	-.020	.049	.022	.010
Social Processes	.038	-.003	-.048	.122 **	.076	-.015
Family	.047	.023	.068	.033	.015	.028
Friends	.063	.006	.027	.044	.037	.080
Humans	.028	.005	.026	.014	.045	.036
Affective Processes	.066	-.034	-.037	.049	.070	.021
Positive Emotion	.075	-.003	.006	.025	.055	.028
Negative Emotion	.002	-.045	-.085 *	.033	.078	-.026
Anxiety	-.038	-.049	-.031	.084	.078	-.007
Anger	.000	-.008	-.046	.059	.081	-.039
Sadness	.004	-.014	-.071	.022	.050	.030

** 2 tailed significance at .01 level
* 2 tailed significance at .05 level

Table C1: Spearman's correlation matrix of the Big Five, privacy concerns and linguistic analysis of Facebook biographies, continued.

	Ex	Ag	Co	Ne	Op	Pr
Cognitive Processes	.038	-.051	-.050	.072	.146 **	-.061
Insight	-.008	-.011	-.004	.037	.118 **	-.030
Causation	.007	-.024	.037	.016	.059	-.020
Discrepancy	-.011	-.068	-.028	.110 *	.070	.070
Tentative	.011	-.076	-.036	.095 *	.102 *	-.064
Certainty	.080	-.004	.053	.065	.039	-.023
Inhibition	-.022	-.001	-.037	.064	.072	.021
Inclusive	-.029	-.104 *	-.028	.060	.046	-.026
Exclusive	-.022	-.031	-.035	.085 *	.123 **	-.041
Perceptual Processes	.084	.031	.016	.027	.034	-.063
See	.084	.007	.073	.002	.047	-.002
Hear	.005	.018	-.031	.079	.065	-.041
Feel	.013	-.018	-.053	.003	-.005	-.071
Biological Processes	.020	.020	-.019	.029	.105 *	.010
Body	-.006	.003	.035	.027	.076	.028
Health	-.007	-.003	-.053	.022	.076	.025
Sexual	.042	.007	.023	.069	.084	.037
Ingestion	.026	-.075	.036	.044	.107 *	-.042
Relativity	.021	-.026	-.030	.066	.066	-.027
Motion	.054	-.018	.055	.018	.049	.030
Space	-.037	-.035	-.025	.047	.083	-.039
Time	.065	-.006	-.017	.045	.066	-.032
Work	-.022	-.110 *	-.047	.049	.049	-.041
Achievement	-.037	-.060	.005	.034	.099 *	.004
Leisure	.060	.001	-.036	.038	.143 **	.009
Home	.002	.033	-.002	-.026	.067	.026
Money	-.034	-.033	.029	-.048	.027	-.032
Religion	.009	.016	.044	.065	.076	-.023
Death	-.005	-.014	.001	-.027	-.014	-.060
Assent	-.001	-.063	-.028	.075	-.003	.021
Nonfluencies	.010	-.036	-.031	.041	-.017	.058
Fillers	.011	-.003	-.063	.062	.022	-.041
Period	.022	.003	-.098 *	.114 **	.111 *	-.018
Comma	-.027	-.054	-.050	.030	.111 *	-.024
Colon	.004	.076	-.028	-.048	.051	-.008
Semicolon	-.003	-.021	.009	-.020	.095 *	.057
Question Mark	-.028	-.067	-.055	.036	-.039	-.019
Exclamation Mark	.048	.016	.001	.010	.015	.012
Hyphen/Dash	-.055	-.035	-.003	-.026	.056	.033
Quotation Mark	.027	.017	-.068	.073	.115 **	.018
Apostrophe	-.031	-.011	-.082	.080	.090 *	-.037
Parentheses	-.065	.025	-.028	.044	.035	.000
Other Punctuation	-.031	.026	-.050	.028	.042	.066
All Punctuation	-.008	.020	-.126 **	.079	.124 **	-.002

** 2 tailed significance at .01 level
* 2 tailed significance at .05 level

Annex D

Table D1: Spearman's correlation matrix of the Big Five, privacy concerns and linguistic analysis of Facebook wall posts.

	Ex	Ag	Co	Ne	Op	Pr
Word Count	.063	.010	-.017	.114 **	.151 **	.002
WPS	.016	.138 **	.058	-.009	-.087 *	.024
Words > Six Letters	-.021	-.046	-.072	-.076	.174 **	.045
Dictionary Words	.018	.054	.129 **	.039	-.049	-.013
Total Function Words	.025	.049	.101 *	.004	.024	.006
Total Pronouns	-.016	-.014	.016	-.003	.043	-.040
Personal Pronouns	.030	-.026	.009	.011	.013	-.014
1st Person Singular	.022	-.063	-.024	.023	.063	-.053
1st Person Plural	.058	.010	.085	-.037	.098 *	.042
2nd Person	-.001	.013	.043	.001	-.002	.007
3rd Person Singular	.095 *	.042	.038	.091 *	-.070	.012
3rd Person Plural	.015	.057	-.052	.047	.081	.030
Impersonal Pronouns	-.090 *	.031	.021	.008	.108 *	-.052
Articles	.076	.067	.149 **	-.068	.057	.014
Common Verbs	.017	.029	.050	.061	-.090 *	-.018
Auxiliary Verbs	.046	.067	.058	.045	-.029	-.024
Past Tense	-.026	-.049	.046	.089 *	-.015	-.037
Present Tense	.072	.065	.074	.046	-.057	-.065
Future Tense	.098 *	.003	-.002	.064	.071	.023
Adverbs	.047	.067	-.010	.078	.074	.006
Prepositions	.095 *	.036	.119 **	-.014	-.019	-.037
Conjunctions	.031	.038	.079	.063	.108 *	.023
Negations	-.047	-.040	-.068	.071	.033	-.028
Quantifiers	.017	.058	.082	.016	.065	.029
Numbers	.011	-.021	.024	.068	.130 **	.060
Swear Words	-.006	-.064	-.107 *	.119 **	.079	-.036
Social Processes	.054	.037	.073	.011	-.025	-.004
Family	.066	.042	.119 **	.013	-.128 **	-.042
Friends	.117 **	.038	.054	.025	.015	-.023
Humans	.051	.042	.049	.070	.037	.022
Affective Processes	.055	.047	.072	.082	-.043	-.048
Positive Emotion	.122 **	.090 *	.161 **	.025	-.076	-.022
Negative Emotion	-.061	-.077	-.110 *	.141 **	.089 *	-.050
Anxiety	.000	.001	.006	.115 **	.016	.002
Anger	-.068	-.058	-.139 **	.098 *	.116 **	-.068
Sadness	.018	-.006	.017	.085 *	.037	.017

** 2 tailed significance at .01 level
* 2 tailed significance at .05 level

Table D1: Spearman's correlation matrix of the Big Five, privacy concerns and linguistic analysis of Facebook wall posts, continued.

	Ex	Ag	Co	Ne	Op	Pr
Cognitive Processes	-.029	.008	.027	.038	.088 *	.016
Insight	.049	-.003	-.036	.047	.053	.008
Causation	-.015	-.015	-.018	.024	.164 **	-.049
Discrepancy	.051	.032	.047	.094 *	.046	-.016
Tentative	-.067	.025	-.048	.055	.094 *	.029
Certainty	.035	.028	.040	-.001	.155 **	.027
Inhibition	-.026	.005	.035	-.002	.035	-.011
Inclusive	.081	.041	.148 **	-.016	.044	-.025
Exclusive	-.038	.008	.027	.062	.065	.020
Perceptual Processes	.031	.039	-.024	.053	.093 *	-.039
See	.085 *	.071	.021	.012	.050	.015
Hear	.013	-.026	-.019	.067	.139 **	-.026
Feel	.034	.021	.062	.112 **	.040	-.008
Biological Processes	.095 *	.022	.003	.162 **	-.056	-.056
Body	.056	.023	-.021	.111 *	-.028	-.048
Health	.015	-.025	.015	.151 **	-.008	-.020
Sexual	.064	.045	-.025	.074	.014	-.076
Ingestion	.139 **	.039	.048	.074	.000	-.035
Relativity	.087 *	.088 *	.174 **	-.013	-.059	-.025
Motion	.066	.003	.166 **	-.036	-.063	-.006
Space	.090 *	.090 *	.116 **	-.055	.035	-.025
Time	.075	.079	.120 **	.057	-.052	-.055
Work	-.039	-.050	-.053	.000	.123 **	-.055
Achievement	.076	.013	.065	-.031	.088 *	-.021
Leisure	.092 *	-.011	.029	.017	.089 *	-.067
Home	.075	.068	.102 *	.094 *	-.034	-.042
Money	.048	-.061	-.056	.029	.111 *	-.023
Religion	-.041	.038	.030	.014	.152 **	-.014
Death	-.001	-.031	-.132 **	.040	.173 **	-.026
Assent	.112 **	.020	-.053	.072	.059	-.025
Nonfluencies	.021	.097 *	-.035	.029	.079	.026
Fillers	.014	.012	-.028	.029	.081	-.004
Period	.065	-.060	.015	-.024	.166 **	.044
Comma	.000	-.006	.048	.010	.104 *	.060
Colon	.007	.046	-.024	-.026	.016	-.033
Semicolon	.089 *	.002	.045	.026	.065	.054
Question Mark	.072	.062	-.106 *	-.022	.092 *	-.078
Exclamation Mark	.166 **	.155 **	.102 *	-.016	-.082	-.046
Hyphen/Dash	-.022	-.013	-.063	.011	.070	.025
Quotation Mark	-.002	.020	.006	-.046	.195 **	-.008
Apostrophe	-.070	.056	-.102 *	.045	.077	.013
Parentheses	.030	.039	-.037	.149 **	.065	.002
Other Punctuation	-.021	-.017	-.104 *	-.012	.120 **	.001
All Punctuation	.060	.045	-.022	-.066	.126 **	.013

** 2 tailed significance at .01 level
* 2 tailed significance at .05 level

Annex E

Table E1: Spearman's correlation matrix of the Big Five, privacy concerns and linguistic analysis of Facebook photo comments.

	Ex	Ag	Co	Ne	Op	Pr
Word Count	.039	-.090	.026	-.018	.107	.082
WPS	.016	-.068	.017	-.011	-.009	.066
Words > Six Letters	-.013	-.056	.052	.010	-.057	.038
Dictionary Words	-.091	-.110 *	-.009	.056	.057	.093
Total Function Words	-.132 *	-.134 *	-.035	.057	.062	.047
Total Pronouns	-.129 *	-.184 **	-.080	.085	.063	.064
Personal Pronouns	-.066	-.166 **	-.031	.049	.035	.067
1st Person Singular	-.048	-.160 **	-.030	.030	.029	.006
1st Person Plural	.054	-.105	.054	-.036	.038	.058
2nd Person	-.007	-.071	.028	.008	-.001	.092
3rd Person Singular	.052	-.059	-.014	.075	.077	.025
3rd Person Plural	.102	-.065	-.030	-.100	.061	.023
Impersonal Pronouns	-.106	-.103	-.113 *	.091	.120 *	.052
Articles	-.035	-.058	-.021	.016	.092	.038
Common Verbs	-.050	-.193 **	-.090	.100	.151 **	.007
Auxiliary Verbs	-.069	-.133 *	-.093	.095	.123 *	.018
Past Tense	-.070	-.085	-.006	.042	.096	.012
Present Tense	-.048	-.137 *	-.104	.066	.127 *	-.004
Future Tense	.002	-.089	-.031	.049	.092	.117 *
Adverbs	-.067	-.136 *	-.062	.030	.071	.019
Prepositions	-.056	-.047	.051	-.013	.028	.068
Conjunctions	.042	-.081	-.002	.044	.057	.050
Negations	.003	-.030	-.065	-.008	.075	.083
Quantifiers	-.012	-.068	.034	-.032	.106	.046
Numbers	.022	-.029	-.003	.014	.061	.038
Swear Words	.032	-.070	-.088	.020	.072	.034
Social Processes	.048	-.078	.020	.034	.085	.123 *
Family	.157 **	.069	.169 **	-.070	.088	.059
Friends	.060	-.072	.001	-.070	.065	.096
Humans	.044	-.083	.040	-.049	.027	.082
Affective Processes	.030	-.059	.002	-.006	.102	.007
Positive Emotion	.049	-.039	.026	-.027	.101	.029
Negative Emotion	-.012	-.127 *	-.042	.019	.092	.001
Anxiety	.016	-.025	-.011	-.006	.006	.034
Anger	.041	-.095	-.033	-.024	.091	-.012
Sadness	.026	-.051	.039	-.009	.051	.013

** 2 tailed significance at .01 level
* 2 tailed significance at .05 level

	Ex	Ag	Co	Ne	Op	Pr
Cognitive Processes	-.049	-.081	-.072	.050	.111 *	.079
Insight	.014	-.020	-.040	.037	.027	.067
Causation	.011	-.021	-.034	-.034	.117 *	.042
Discrepancy	.019	-.084	-.033	-.023	.096	.064
Tentative	-.064	-.072	-.053	.073	.117 *	.002
Certainty	.029	-.096	-.080	-.017	.127 *	.042
Inhibition	.043	-.007	-.011	-.095	.094	.074
Inclusive	.027	-.096	-.013	-.007	.064	.059
Exclusive	-.049	-.013	.004	-.017	.082	.042
Perceptual Processes	.020	-.040	.046	-.002	.072	.062
See	.004	-.044	.039	-.064	.061	.017
Hear	.061	-.144 **	.058	.008	.088	.066
Feel	.040	-.058	.011	.013	.106	.099
Biological Processes	.048	-.127 *	.068	.033	.043	.040
Body	.016	-.098	.023	.014	.015	.030
Health	-.013	-.078	.058	-.089	-.009	.025
Sexual	.069	-.123 *	.065	-.012	.054	.064
Ingestion	.083	-.082	.056	-.049	.047	.006
Relativity	-.046	-.083	.061	.012	.059	.109
Motion	-.031	-.115 *	-.039	.010	.023	.106
Space	.028	-.036	.061	-.073	.069	.065
Time	-.073	-.075	.008	.046	.057	.145 **
Work	.006	.014	-.002	-.025	.091	.041
Achievement	-.011	-.112 *	-.017	.001	.044	.091
Leisure	-.003	-.125 *	-.010	.000	.095	.091
Home	.067	-.028	.002	.015	.123 *	.116 *
Money	.086	-.097	-.060	-.052	.088	.076
Religion	.058	.001	.055	-.082	.007	-.048
Death	.035	-.019	-.034	.043	.146 **	-.071
Assent	.052	-.007	-.031	-.030	.046	-.012
Nonfluencies	.012	.007	-.091	-.091	.018	.030
Fillers	.046	.013	-.028	-.030	.049	.045
Period	.015	-.070	-.006	.068	.071	.003
Comma	-.036	-.081	.021	.018	.038	.056
Colon	-.014	-.025	-.035	.005	-.041	.018
Semicolon	.009	-.082	-.029	.048	.065	.057
Question Mark	.061	-.020	.019	-.049	.049	.049
Exclamation Mark	.019	-.020	.055	-.021	.068	-.013
Hyphen/Dash	.033	-.033	.042	-.030	.073	.083
Quotation Mark	.052	-.052	-.068	.027	.215 **	-.012
Apostrophe	-.039	-.104	-.069	.025	.034	.045
Parentheses	.038	.002	.061	-.065	.191 **	.077
Other Punctuation	.037	-.024	-.008	.016	.006	.083
All Punctuation	.020	-.060	-.034	.051	.052	.014

** 2 tailed significance at .01 level
* 2 tailed significance at .05 level