Predicting Individual Characteristics from Digital Traces on Social Media: A Meta-Analysis

Title: Predicting Individual Characteristic from Digital Traces on Social Media: A Systematic Review and Meta-Analysis

Michele Settanni[a], Danny Azucar[a]* & Davide Marengo[a]

*Corresponding author: Danny Azucar, dazucarc@unito.it

[a] Department of Psychology, University of Turin, 10124 Via Verdi 10, Turin, Italy

Key words: Social Media, Digital Traces, Psychosocial Characteristics, Psychological Assessment, Data Mining, Predictive Modeling

Abstract

The increasing utilization of social media provides a vast and new source of user generated ecological data (digital traces), which can be automatically collected for research purposes. The availability of these datasets, combined with the convergence between social and computer sciences, has led researchers to develop automated methods to extract digital traces from social media and use them to predict individual psychological characteristics and behaviors. In this paper, we reviewed the literature on this topic and conducted a series of meta-analyses to determine the strength of associations between digital traces and specific individual characteristics; personality, psychological well-being, and intelligence. Potential moderator effects were analyzed with respect to type of social-media platform, type of digital traces examined, and study quality. Our findings indicate that digital traces from social media can be studied to assess and predict theoretically distant psychosocial characteristics with remarkable accuracy. Analysis of moderators indicated that the collection of specific types of information (i.e., user demographics), and the inclusion of different types of digital traces, could help improve the accuracy of predictions.

Introduction

1.1 Emergence of Social Media

The recent years have seen a major evolution in how people interact with each other through the Internet, [1] and the growth of social network sites and social media has yielded great sources of online interpersonal communication, with users spontaneously expressing themselves in a naturalistic setting about everyday topics and events. [2-6] This ever-increasing utilization of social media provides a vast and new source of user generated ecological data with connections to offline personal characteristics, attitudes, and behaviors. [5-11] Digitally mediated user behaviors on social media, along with the information individuals share on their personal profiles, are recorded and have been collected and analyzed by researchers from diverse disciplines. More specifically, due to the popularity of social media, psychologists have begun studying the relationships between psychosocial characteristics and digitally mediated human behaviors, or '*digital traces*'. [6, 12-14] In this field of research, the terms 'digital traces', 'digital footprints', and 'digital records' are used interchangeably; throughout this paper we use the term 'digital traces' for consistency. Defined as information generated by users on their social media profiles, digital traces consist of personal information about age, gender, sexual orientation, and location, as well as activity information including network size, shared text, pictures, and videos.[15] Access to these data sources is changing the way researchers use content analysis to understand people, effectively opening the gate for the collaboration between the social and computer sciences.[6] The increasing availability of large datasets from social media, fostered by this convergence of disciplines, has allowed researchers to not only seek to *gain insights* from studying human behaviors on social media, but also to *predict*

psychological characteristics and behaviors based on automated data mining and the analysis of digital traces. Studies using automated approaches are mostly aimed at developing models to predict individual characteristics using all the vast information available on social media profiles (e.g., predicting individual personality using data referring to activity statistics, language use, and pictures posted on social media). [6, 14] Unlike traditional explanatory studies, this type of researches are mainly data-driven, using features extracted from digital traces to predict psychological characteristics without referring to specific *a priori* theories or hypotheses.[6, 16] This kind of predictive research does not require causality assumptions concerning the relationship between digital traces and predicted outcomes. Instead, the focus is on association rather than causation, meaning that there is no need to provide explanation of the exact role of each variable in terms of an underlying causal structure, and criteria for choosing predictors mostly relates to strength of the association between the predictors and the outcomes, and availability and quality of collected data.[17] The present review focuses on this type of researches.

1.2 Predicting individual characteristics via automated analysis of digital traces

Studies focusing on the prediction of psychosocial and behavioral characteristics based on digital traces from social media generally use a common methodology, consisting of the following steps: *(1)* users are contacted and asked to complete self-report questionnaires assessing the characteristic of interest, and provide complete or limited access to their digital traces on social media, *(2)* digital traces are collected and analyzed using automated approaches to extract sets of profile attributes, or *features (e.g., activity statistics,* such as number of friends, and status updates; *linguistic features*, such as frequency of words in pre-

defined categories in posts), and *(3)* the predictive power of these features is examined over participants' individual characteristics as assessed via self-reports, using a varied set of predictive methods, ranging from univariate linear regression modeling to classification via machine learning. One of the earliest and largest projects employing this approach is the MyPersonality project, [13] which has cultivated a dataset consisting of self-report data for a wide range of behavioral and psychosocial characteristics (e.g., personality, satisfaction with life, substance use), and digital traces of over 70,000 Facebook users. Many researchers have used this data set to conduct automatic coding of user profiles and assess or predict distinguishing features of user personality and well-being. [5,6,9,10,13,18-22] Furthermore, scholars have demonstrated the feasibility of predicting many psychosocial characteristics from features extracted from a variety of digital traces (e.g., user demographics, [20] user activity statistics, [8, 18] linguistic features, [6] and features extracted from pictures [23], social media platforms (e.g., Facebook, [24] Twitter, [25] Sina Weibo [2]), and by employing different analytical approaches (e.g., use of a single type of digital trace [8] vs. multiple sources of digital traces [20]). However, due to the existing heterogeneity among researches of this young field of research, there is a need to synthesize and summarize the existing literature in order to evaluate their accuracy and recommend the best methods to predict psychological characteristics based on the analysis of digital traces collected from social media.

1.3 Aims

The aim of the current study is to conduct a series of meta-analyses to determine the mean effect-size of associations between digital traces from social media and specific individual characteristics. Meta-analyses were conducted on characteristics investigated by at least three

studies, namely personality, psychological well-being, and intelligence. Given the expected presence of effect size heterogeneity among studies, potential moderator effects were analyzed with respect to the following variables: type of social-media platform (public vs. private), type of digital traces examined, and study quality.

2. Methods

2.1. Search strategy and inclusion criteria

An initial dataset of 1,677 articles was identified during the month of July 2016 by submitting a search query to the Scopus, ISI Web of Science, PubMed, and Proquest databases. The query searched keywords in the 'title', 'abstract' and 'keyword heading' fields, where available. The following keywords and stems were used in both separate and combined searches:

psych, behavior, personality, health, well being, risk, depression, quality of life, life satisfaction, risk behavior, substance, abuse, psychological assessment, cyber psychology, emotional well being, mental health, gender, age,* in conjunction with *myspace, facebook, instagram, twitter, youtube, photobucket, linkedin, social network, reddit, social media, snapchat, periscope, social networking, status updates, mypersonality, machine learning, data mining, text analysis, language processing, closed vocabulary, closed dictionary, liwc, open vocabulary, open dictionary, support vector machines, text mining, topic modeling, dictionary, latent dirichlet allocation, differential language analysis, digital footprint, differential language, computational social science, content analysis, linguistic studies*

After duplicates were removed, a set of 1,241 articles was screened for the following inclusion criteria - 1. Studies must focus on human behavioral or psychological characteristics, 2. Studies must focus on *individual* human behavioral or psychological characteristics, 3. Studies

must be at least partially quantitative in nature, 4. Studies must analyze digital traces of human behavior, and 5. Studies must include a valid self-report measure to assess individual characteristics. A total of 1,203 articles were excluded upon inspection of their abstracts, and full text assessment for eligibility was conducted for 38 articles. This screening process resulted in the initial selection of twenty-five (*n* = 25) articles for inclusion in our analysis. We then identified an additional thirty-four (*n* = 34) articles through a review of the 'citations' from the twenty-five originally selected articles, and of these, thirteen (*n* = 13) were selected for inclusion in our review based on the aforementioned inclusion criteria. This resulted in a final set of 38 articles selected for the review. The flow-chart of the article selection process is depicted in Fig. 1.

2.2 Research coding

Coding of psychological and behavioral characteristics. Investigated characteristics varied across studies. We identified three general psychological characteristics, which were investigated at least by 3 studies: personality traits (Big 5 and Dark Triad), psychological well-being (depression, anxiety and stress, life satisfaction), and intelligence. Other characteristics were present in less than 3 studies: personal values, coping strategies, substance use, and self-monitoring skills (see Table 1).

Coding of digital traces. Studies varied considerably in terms of digital traces analyzed. We distinguished between the following types of digital traces: (1) User demographics (e.g., gender, age, location), (2) User activity statistics (e.g., number of posts, number of contacts or friends, number of received Likes, comments, mentions), (3) Language (e.g., Twitter's tweets, Facebook's status updates and comments), (4) Facebook Likes (i.e., expression of interest in Facebook pages about events, persons, locations, products, etc.), and (5) Pictures (e.g., profile

pictures, Instagram photos).

Coding of moderators. Studies differ considerably with regard to the type of social media platform, and type of examined digital traces.

Concerning the distinction between types of social media platform, we chose to group social-media based on their default privacy settings, distinguishing between public (social media platforms that make posts and updates public by default, i.e, Twitter, Sina Weibo, Reddit, Instagram) and private (social media platform in which user posts are visible only by users' friends, i.e., Facebook)

These factors may influence the magnitude of the accuracy of prediction of psychological characteristics. We considered six potential moderators, that were dichotomously coded: (1) Type of social-media platform (Private vs. Public), (2) Use of user demographics (yes vs. no), (3) Use of activity statistics (yes vs. no), (4) Use of language-based features (yes vs. no), (5) Use of pictures (yes vs. no), (6) Use of multiple vs. single types of digital traces (e.g., language vs. language+pictures). We also added a (7) study quality moderator. Given that the heterogeneity of the research areas in which analyzed studies were conducted makes it impossible to define a methodological standard, study quality was assessed using the quality of the source the study was published in. Papers were categorized into top, middle, and low tiers using the quartile that sources belong to in the 2016 Scopus CiteScore; ranking quartile 1 as top tier (high quality), quartile 2 as middle tier (medium quality), and quartile 3 or 4 and non-indexed studies as low tier (low quality).

Independence of studies. When selecting studies for inclusion in the meta-analyses, we

found that several articles contained potentially overlapping samples. In particular, studies using data collected by the MyPersonality project, for example, potentially share parts of the same sample and data, and often investigate the same main characteristic. In general, a potential lack of independence exists between many of those studies, violating certain statistical assumptions of the meta-analysis. In efforts to resolve this issue, we followed recommendations from previous studies. [26, 27] We considered studies as non-independent if they met the following criteria: (1) each correlation was based on responses from overlapping sample subjects, (2) the main assessed characteristics were the same, (3) digital traces were extracted from the same social media platform, and (4) type of digital traces used to predict characteristics were the same or partly overlapping. When studies were found to be non-independent based on the aforementioned criteria, the paper with the most comprehensive set of digital traces was included in the analysis. In the case of non-independent studies analyzing the same set of digital traces, the one with the larger sample size was included in the meta-analysis. In the case of studies including more than one effect-size referring to the same psychological characteristic (e.g., Big 5 traits), we averaged the effect-sizes to obtain a single effect size to ensure independence of the correlations entered into the meta-analysis. [28]

2.3 Strategy of Analyses

For each study, an effect size was calculated. We used Pearson's r to express the relationship between digital traces and investigated outcomes. We chose not to transform correlations into Fisher's z scores for meta-analytic calculations because this transformation produces an upward bias in the estimation of mean correlation, which is usually higher than the downward bias produced by the use of untransformed correlations. [29]

When studies did not report Pearson's r, but instead reported alternative effect-size indicators (e.g., when characteristics were examined in dichotomous form by distinguishing individuals at low and high levels using validated or empirically derived cut-offs), reported effect-sizes were converted to correlations. Area Under the Receiver Operating Characteristic curve (AUROC) statistics were first converted to Cohen's d [30], and then converted from Cohen's d to *r.* [31] When studies provided specificity and sensitivity values, or positive predicted values (PPV) and negative predicted values (NPV), or enough information was available for computing these statistics, we used this information to compute odds-ratios, [32] then transformed odds-ratios to Cohen's d, [33] and finally converted Cohen's d to correlations. [31] When studies only reported the mean absolute error (MAE) and root mean square error (RMSE) statistics (*n* = 7), and thus did not provide enough information to compute correlations, or results were not fully reported in the study (*n* = 2), we contacted the first author of the study to obtain any missing information. Missing information was obtained for one study (*n* = 1).

We conducted separate meta-analyses for each main characteristic (i.e., personality, psychological well-being, and intelligence). Meta-analyses were performed using a random-effects model as the true effect size was likely to vary in the individual studies; owing to the variety in data sources, study designs, and analytic approaches. Grubb's test was used to identify outliers. Heterogeneity of the studies' effect-sizes included in each pooled analysis was evaluated by examination of (1) the chi-square Q statistic of heterogeneity, (2) the T^2 estimate of true between-study variance, and (3) the I^2 statistic of proportion of variation in observed effects due to the variation in true effects. Possible publication bias was evaluated by inspecting the funnel plot, by the statistical significance of the Begg and Mazumdar adjusted rank

correlation test [34] and Egger's test of the intercept, [35] Duval and Tweedie's trim and fill procedure, [36] and classic fail-safe *N*.

Then, potential moderators were analyzed using meta-regression models. The effect of moderators on study effect-sizes was measured by random-effects univariate meta-regressions using maximum-likelihood estimation. For the purpose of moderator analyses, and in order to obtain sufficiently robust coefficient estimates, we followed the suggestion by Fu and colleagues [37] and examined the effect of grouping variables only if at least 4 studies per group were available. We employed a critical value of $\alpha = .05$ in our meta-regression analyses, but due to the low number of studies, effects approaching statistical significance ($p < .10$) are commented as suggestions of possible links which are worthy of being explored by future researches.

3. Results

3.1 Overview of studies

We found 38 papers, resulting in 50 different effect-sizes (see Table 1). Information about all selected studies is shown in Tables 1-2.

Overall, we found three characteristics for which at least three studies were published, namely personality (26 papers [1, 4, 6-10, 13, 14, 18-22, 25, 38-48] including 30 effect-sizes), psychological well-being (10 papers [2, 3, 5, 11, 13, 19, 24, 49-51] including 10 effect-sizes), and intelligence (3 papers [13, 19, 23] including 3 effect-sizes). Other characteristics for which we found fewer than three studies were: social satisfaction, [2] substance use, [13] self-monitoring skills, [21, 52] personal values, [38, 53] and coping style. [54]

Meta-analyses were performed on characteristics that were reported in at least three studies. After inspection of studies for non-independence, we selected a subset of 25 papers

including 30 independent effect-sizes about the three main characteristics, namely personality (n = 18), psychological well-being (n = 9), and intelligence (n = 3) (see Table 1). Grubb's test failed to identify any outliers, resulting in no further studies being excluded. Results of meta-analyses are reported below.

3.2 Meta-analyses

3.2.1 Personality

Mean effect size. To establish the magnitude of the association between digital traces and personality, we analyzed 18 independent effect-sizes. The estimated meta-analytic correlation was .34, 95% CI [.27 - .34] (Fig. 2), and this effect was significantly greater than zero, z = 9.58, p < .001. Q test for heterogeneity was significant: Q (17) = 318.33 (p < .001). There was low true heterogeneity between studies, T^2 = 0.02 (T = 0.14), and the observed dispersion of effect-sizes was mostly due to true heterogeneity (I^2 = 94.66).

Publication bias. First, we inspected the funnel plot (Fig. 3), plotting the included studies' effect size against its standard error. The funnel plot was symmetrical, suggesting lack of publication bias. Trim-and-fill analysis suggested that no studies were missing on the left side of the mean effect. The p values of Begg and Mazumdar test and Egger's test were p = 0.52 and p = 0.43, indicating no significant evidence of publication bias. The result of classic fail-safe N suggested that 9638 null reports would be required in order for the combined 2-tailed p-value to exceed the alpha level of .05. The fail-safe N value was larger than 100, corresponding to the recommended rule-of-thumb limit of 5k+10.[55] The results of these four tests indicated that it is unlikely that publication bias poses a significant threat to the validity of the findings reported in the current analysis.

Moderator analyses. We examined the following moderating effects: (1) Privacy vs. public oriented social-media platform, (2) Multiple vs. single types of digital traces, (3) Use of user demographics, (4) Use of activity statistics, (5) Use of language-based features, (6) Use of pictures, (7) Study quality. Results of univariate meta-regressions, shown in Table 3, indicated an increase in strength of association between digital traces and personality when studies examined multiple types of digital traces compared with only one type (K = 18, β = 0.19, p<.05). Use of demographic statistics for prediction purposes was also associated with an increase in correlation strength between digital traces and personality (K = 18, β = 0.23, p <.05). The remaining moderators did not show significant effects.

3.2.2 Psychological Well-being

Mean effect size. The magnitude of the association between digital traces and psychological well-being was analyzed by summarizing 9 independent effect sizes. The estimated meta-analytic correlation was .37, 95% CI[.28 - .45] (Fig.5), and this effect was significantly greater than zero, z = 7.54 , p < .001. Q test for heterogeneity was significant: Q (8) = 124.67 (p < .001). There was relatively low true heterogeneity between studies, T^2 = 0.02 (T = 0.14), and the observed dispersion of effect-sizes was mostly due to true heterogeneity (I^2 = 93.58).

Publication bias. Inspection of funnel plot (Fig. 5), and trim-and-fill analysis suggested that no studies were missing on the left side of the mean effect. The p values of Begg and Mazumdar test and Egger's test were p = 0.37 and p = 0.07, indicating low probability of publication bias. The result of classic fail-safe N suggested that 1618 null reports would be required in order for the combined 2-tailed p-value to exceed the alpha level of .05. The fail-safe N value was larger

than the recommended rule-of-thumb limit of 55. Overall, results did not suggest existence of significant publication bias.

Moderator analyses. We examined the following moderating effects: (1) Multiple vs. single sources of digital traces (2) Type of social media platform (Private vs. Public) (3) Use of activity statistics (4) Study quality. Remaining categorical moderators were not tested because they did not reach the per-group minimum value of 4 distinct studies. Results of univariate meta-regressions (Table 3) seem to indicate a relevant increase in the effect size of the association between digital traces and psychological well-being when using multiple types of digital traces compared with use of only one type ($K = 9$, $\beta = 0.18$, $p < .10$). Additionally, when comparing studies conducted on private social media platform (e.g., Facebook) with those conducted on public platforms (e.g., Twitter), a relevant difference in effect size in favor of public platforms emerged ($\beta = -0.18$, $p < .10$), even if the effect did not reach proper significance. However, given the perfect collinearity between the variables concerning private/public platforms and multiple/single type of digital traces, a univocal interpretation of these moderator effects is not possible. A larger and more differentiated sample of studies will permit to ascertain both the presence of a significant impact of the type of platform, and distinguish between the effects of multiple vs. single types of digital traces and type of platform analyzed. Remaining moderators did not show relevant effect.

3.2.3 Intelligence

Mean effect size. The magnitude of the association between digital traces and intelligence was analyzed by summarizing effects presented in 3 studies. The estimated meta-analytic correlation was .29, 95% CI[.19 - .38] (Fig. 6), and this effect was significantly greater than zero,

z= 5.65 , p < .001. Q test for heterogeneity was significant: Q (2) = 24.01 (p < .001). There was low between-study heterogeneity, T^2 = 0.01 (T = 0.09), and the observed dispersion of effect-sizes was mostly due to true heterogeneity (I^2 = 91.67).

Publication bias. Upon examination, funnel plot (Fig. 7) was found to be symmetrical, suggesting no publication bias. Additionally, trim-and-fill analysis suggested that no studies were missing on the left side of the mean effect. The Begg and Mazumdar's (p = 0.99) and Egger's test (p = 0.84), and the result of classic fail-safe N suggested (463 null reports required to exceed the alpha level of .05) suggested lack of publication bias.

4. Discussion

To our knowledge, this is the first meta-analysis summarizing results from studies investigating the use of digital traces collected from social media to predict psychological and behavioral characteristics. Our main aim was to determine the mean effect-size of associations between digital traces from social media and specific individual characteristics. Based on the review of the literature, we found 38 articles employing features automatically extracted from digital traces of human behavior on social media to predict different psychosocial characteristics. Meta-analyses were conducted on characteristics investigated by at least three independent studies, namely personality, psychological well-being, and intelligence. Overall, we found the majority of reported associations between features extracted from digital traces and investigated characteristics to be at least of moderate strength. Significant associations with digital traces were found for each of the most investigated characteristics, with mean correlation values (Pearson's r) ranging from .29 (intelligence) to .37 (psychological well-being). Included effect-sizes showed low-to-moderate dispersion that was mostly due to true

differences across studies. Given the presence of heterogeneity of effects among studies, potential moderator effects were analyzed with respect to the following possible sources of variation: type of social media platform (private vs. public), type of extracted features, and analytical approaches (e.g., use of multiple vs. single type of digital traces). Our hypothesis was that each of these factors and their interactions could contribute to the overall heterogeneity of effects. Unfortunately, given the small number of studies included in the meta-analyses, we were able to perform moderator analyses for only two characteristics; personality and psychological well-being. Moreover, we were able to investigate the influence of only a subset of the moderators, and it was not possible to test the influence of interaction effects between moderators. Our results indicate that the association between digital traces, and both personality and well-being, was stronger when multiple types of digital traces were analyzed. Regarding the type of extracted features, use of demographics extracted from social media positively affected the strength of the relationship between personality and digital traces, suggesting the opportunity to include them in models aimed at increasing the predictive power of digital traces. Furthermore, the type of social media platform (public vs. private) did not affect the strength of association with personality, while digital traces extracted from private platforms were less strongly associated with psychological well-being.

Overall, analysis of moderators pointed out that a significant part of the effect size heterogeneity can be traced back to the amount of digital traces included in the studies: generally, higher effect sizes have been achieved by studies including multiple types of digital traces. We hypothesize that future studies will confirm this relationship; hence, in order to reach a higher predictive power scholars should collect data from a large set of different digital

traces, possibly combining different types of data (e.g., pictures and text) from different social media platforms.

As noted in the Introduction, most of the reviewed studies focused on predicting individual characteristics without providing explanations or hypotheses regarding the existing relationships between specific digital traces and outcomes. In fact, published studies have mostly focused on developing statistical models and on maximizing predictive accuracy. This approach is quite common in computer science, while still relatively novel among other disciplines. As this approach becomes more common in psychology and social sciences, we expect that findings of predictive studies may significantly contribute to the refinement of existing, and the building of new theories.

4.1 Limitations and Future Directions

The present study has limitations. First, given the low number of included studies per social media platform, we were not able to compare the impact of specific social media platforms on the ability to predict individual characteristics based on digital traces. As social media platforms differ significantly among each other (e.g., in the way users interact with each other, in the type of shared data, privacy settings, and user demographics), we expect that differences may emerge in the strength of association between digital traces and psychological characteristics when comparing different social media. As more studies in this area are conducted, it will be possible to test this hypothesis and determine the specific impact these differences have on prediction accuracy. In the present study, we were only able to compare social media platforms based on their default privacy settings. However, since the majority of social media platforms

allow individual users to customize their specific privacy settings, findings should be taken with caution. Future studies exploring the impact of privacy settings on the use of digital traces for the prediction of psychosocial characteristics should aim to also gather information about user's selected privacy settings. Finally, the present review was unable to investigate the effect cultural differences may have on the associations between digital traces from social media and specific individual characteristics. Almost all of the included articles were based on samples of English-speaking users, while different contexts were largely underrepresented. As more studies will be conducted on more culturally diverse samples, it will be possible to investigate the impact of national/cultural differences may have on the associations between digital traces from social media and specific individual characteristics.

4.2 Ethical Considerations

It is worth to note that most of the studies analyzed in this review do not carefully address the ethical issues arising from developing techniques to assess individual characteristics on the basis of data recorded and collected from social media. The possibility to identify individuals with specific characteristics or to screen large samples of individuals for certain behaviors represents beneficial opportunities, given the users' consent, to target public health interventions and for customized advertising. However, possible misuses (or questionable uses) of these tools exist: recently, newspapers reported cases showing the feasibility and the efficacy of targeting political messages on the basis of unintentionally disclosed information on social media [56, 57] or targeting ads on the basis of users' emotional state. [58] The risks associated with the application of these new techniques to specific areas and subjects should be carefully considered by scholars. In order to protect individuals' privacy and prevent abusive behaviors it would be desirable that

awareness about these issues spread among both policymakers and public audience.

4.3 Conclusions

The present meta-analysis demonstrates that digital traces extracted from social media can be used to infer specific psychological characteristics. The presence of significant associations between digital traces and psychosocial characteristics, and the lack of relevant differences in the strength of these associations, indicate that records of digitally mediated behaviors from social media can be used to study and predict theoretically distant psychosocial characteristics with comparable accuracy. ~~This result is encouraging for scholars who aim to use data extracted from social media to study different and not yet investigated characteristics.~~ The relationship between digital traces of online behavior and psychological characteristics is quite strong, apparently stronger than the association found by scholars studying the link between personality and offline behaviors. [59, 60] Given the relative novelty of this approach and the fast technological evolutions that make it possible to access the ever-growing sources of online data, the accuracy in the prediction of individual characteristics is expected to improve steeply in the next few years. We expect the accuracy to grow because of the ongoing transition from the use of traditional analytic approaches toward a more pervasive employment of data mining techniques (e.g., machine learning algorithms [61]), as well as the emergence of new techniques to extract meaningful information from visual data (i.e., image recognition via artificial intelligence [62]), which is especially important, given the current shift in content sharing on social media, from text, to photos and videos. [63] These methodological improvements will hopefully help this research area become more mainstream among social scientists, which in turn will favor the theoretical reflection regarding the relationship between actual online behaviors and

individual characteristics.

In conclusion, findings from the present study indicate that digital traces of human behavior from social media represent a relevant source of information for the prediction of individual psychological characteristics as diverse as personality, intelligence, and psychological well-being. Moreover, we show that the collection of specific types of information (e.g., user demographics), and the inclusion of different types of digital traces, can help improve the accuracy of these predictions. These results have implications on the development of tools allowing for the unobtrusive assessment of psychological characteristics of social media users, which in turn can be beneficial for a variety of purposes, including commercial applications (e.g., user-tailored advertising and online experiences) and health-related purposes (e.g., early detection of individuals at risk for depression, longitudinal tracking of mental well-being trends).

Acknowledgments

No competing financial interests exist.

References

1. Garcia D, Sikström S. The dark side of Facebook: Semantic representations of status updates predict the Dark Triad of personality. Personality and Individual Differences 2014; 67: 92-96. DOI: 10.1016/j.paid.2013.10.001

2. Bai S, Gao R, Hao B, Yuan S, Zhu T. (2014). Identifying Social Satisfaction from Social Media. arXiv preprint arXiv:1407.3552.

3. De Choudhury M, Gamon M, Counts S, Horvitz E. 2013. Predicting depression via social media. Proc. 7th Int. AAAI Conf. Weblogs Soc. Media, Boston, MA, July 8–July 10

4. Farnadi G, Sitaraman G, Sushmita S, et al. Computational personality recognition in social media. User Modeling and User-Adapted Interaction 2016; 1-34. DOI: 10.1007/s11257-016-9171-0

5. Liu P, Tov W, Kosinski M, Stillwell DJ, Qiu L. Do Facebook Status Updates Reflect Subjective Well-Being? Cyberpsychology, Behavior, and Social Networking 2015; 18(7): 373-379. DOI: 10.1089/cyber.2015.0022

6. Schwartz HA, Eichstaedt JC, Kern ML, et al. Personality, gender, and age in the language of social media: The open-vocabulary approach. PloS one 2013; 8(9): e73791. DOI: 10.1371/journal.pone.0073791

7. Golbeck, J., Robles, C., & Turner, K. Predicting personality with social media. In CHI'11 extended abstracts on human factors in computing systems, 2011; pp. 253-262). ACM. DOI: 10.1145/1979742.1979614

8. Gosling SD, Augustine, AA, Vazire S, Holtzman N, Gaddis S.Manifestations of personality in online social networks: Self-reported Facebook-related behaviors and observable profile

information. Cyberpsychology, Behavior, and Social Networking 2011; 14(9): 483-488. DOI: 10.1089/cyber.2010.0087

9. Kern ML, Eichstaedt JC, Schwartz HA, et al. The online social self an open vocabulary approach to personality. Assessment 2013; 21(2): 158-169. DOI: 10.1177/1073191113514104

10. Park G, Schwartz HA, Eichstaedt JC, et al. Automatic personality assessment through social media language. Journal of personality and social psychology 2015; 108(6): 934- 952. DOI: 10.1037/pspp0000020

11. Settanni M, Marengo D. Sharing feelings online: studying emotional well-being via automated text analysis of Facebook posts. Frontiers in psychology 2015; 6. DOI: 10.3389/fpsyg.2015.01045

12. Farrell D, Petersen JC. The growth of internet research methods and the reluctant sociologist. Sociological Inquiry 2010; 80(1): 114-125. DOI: 10.1111/j.1475-682X.2009.00318.x

13. Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior. Proceedings of the National Academy of Sciences 2013; 110(15): 5802-5805. DOI: 10.1073/pnas.1218772110

14. Li L, Li A, Hao B, Guan Z, Zhu T. Predicting active users' personality based on micro-blogging behaviors. PloS one 2014; 9(1): e84997. DOI: 10.1371/journal.pone.0084997

15. Madden M., Fox. S., Smith. A., & Vitax. J. (2007) Digital Footprints. *Pew Research Center*. Retrieved from http://www.pewinternet.org/2007/12/16/digital-footprints/

16. Schwartz, H. Andrew & Lyle H. Ungar. "Data-driven content analysis of social media: a systematic overview of automated methods." *The ANNALS of the American Academy of Political and Social Science* 659.1 (2015): 78-94.

17. Shmueli, Galit. "To explain or to predict?." *Statistical science*25.3 (2010): 289-310.

18. Bachrach Y, Kosinski M, Graepel T, Kohli P, Stillwell D. (2012) Personality and patterns of Facebook usage. In Proceedings of the 4th Annual ACM Web Science Conference 2012; (pp. 24-32). ACM. DOI: 10.1145/2380718.2380722

19. Kosinski M, Bachrach Y, Kohli P, Stillwell D, Graepel T. Manifestations of user personality in website choice and behaviour on online social networks. Machine learning 2014; 95(3): 357-380. DOI: 10.1007/s10994-013-5415-y

20. Markovikj D, Gievska S, Kosinski M, Stillwell D. Mining facebook data for predictive personality modeling. In Proceedings of the 7th international AAAI conference on Weblogs and Social Media 2013, Boston, MA, USA.

21. Quercia D, Lambiotte R, Stillwell, D, Kosinski M, Crowcroft J. The personality of popular facebook users. In Proceedings of the ACM 2012 conference on computer supported cooperative work (pp. 955-964). ACM. DOI: 10.1145/2145204.2145346

22. Youyou W, Kosinski M, Stillwell D. Computer-based personality judgments are more accurate than those made by humans. Proceedings of the National Academy of Sciences 2015; 112(4): 1036-1040. DOI: 10.1073/pnas.1418680112

23. Wei X, Stillwell D. (2016). How smart does your profile image look? Intelligence estimation

from social network profile images. arXiv preprint arXiv:1606.09264.

24. Schwartz HA, Eichstaedt J, Kern ML, et al. (2014). Towards assessing changes in degree of depression through facebook. In Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality (pp. 118-125). DOI: 10.3115/v1/W14-3214

25. Liu L, Preotiuc-Pietro D, Samani ZR, Moghaddam ME, Ungar LH. (2016). Analyzing Personality through Social Media Profile Picture Choice. In ICWSM (pp. 211-220).

26. Hunter J, Schmidt F, Jackson G. (1982). Meta-Analysis: Cumulating research findings across studies. Beverly Hills CA: Sage.

27. Sheppard BH, Hartwick J, Warshaw PR. The theory of reasoned action: A meta-analysis of past research with recommendations for modifications and future research. Journal of consumer research 1988: 15(3), 325-343. DOI: 10.1086/209170

28. Hunter JE, Schmidt FL. Dichotomization of continuous variables: The implications for meta-analysis. Journal of Applied Psychology 1990; 75: 334–349. DOI: 10.1037//0021-9010.75.3.334

29. Schmidt FL, Hunter JE (2014). Methods of meta-analysis: Correcting error and bias in research findings. Sage publications. DOI: 10.4135/9781483398105

30. Ruscio J. A probability-based measure of effect size: Robustness to base rates and other factors. Psychological Methods 2008; 13: 19-30. DOI: 10.1037/1082-989X.13.1.19

31. Rosenthal R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), The Handbook of Research Synthesis. New York, NY: Sage. pp. 239.

32. Glas AS, Lijmer JG, Prins MH, Bonsel GJ, Bossuyt PM. The diagnostic odds ratio: a single

indicator of test performance. Journal of clinical epidemiology 2003; 56(11): 1129-1135. DOI: 10.1016/S0895-4356(03)00177-X

33. Borenstein M, Hedges LV, Higgins J, and Rothstein H.R. Meta-Regression. Introduction to Meta-analysis 2009: 187-203.

34. Begg CB, and Madhuchhanda Mazumdar. Operating characteristics of a rank correlation test for publication bias. Biometrics 1994: 1088-1101.

35. Sterne, JA, Matthias E. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. Journal of clinical epidemiology 54.10 2001: 1046-1055.

36. Duval S, Tweedie R. Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. Biometrics 56.2 2000: 455-463.

37. Fu R, Gartlehner G, Grant M, et al. Conducting Quantitative Synthesis When Comparing Medical Interventions. Methods Guide for Effectiveness and Comparative Effectiveness Reviews 2014: 254.

38. Celli F, Bruni E, Lepri B. Automatic personality and interaction style recognition from facebook profile pictures. Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014. DOI: 10.1145/2647868.2654977

39. Gao R, Hao B, Bai S, et al. Improving user profile with personality traits predicted from social media content. Proceedings of the 7th ACM conference on recommender systems. ACM, 2013. DOI: 10.1145/2507157.2507219

40. Golbeck J. Predicting Personality from Social Media Text. AIS Transactions on Replication Research 2.1 2016: 2.

41. Kleanthous S, Herodotou C, Samaras G, Germanakos P. Detecting Personality Traces in

Users' Social Activity. International Conference on Social Computing and Social Media. Springer International Publishing, 2016. DOI: 10.1007/978-3-319-39910-2_27

42. Preotiuc-Pietro D, Carpenter J, Giorgi S, Ungar L. Studying the Dark Triad of Personality through Twitter Behavior. Proceedings of the 25th ACM International on Conference on Information and Knowledge Management. ACM, 2016. DOI: 10.1145/2983323.2983822

43. Qiu L, Lin H, Ramsay J, Yang F. You are what you tweet: Personality expression and perception on Twitter. Journal of Research in Personality 46.6, 2012: 710-718. DOI: 10.1016/j.jrp.2012.08.008

44. Skowron M, Tkalcic M, Ferwerda B, Schedl M. Fusing social media cues: personality prediction from twitter and instagram. Proceedings of the 25th international conference companion on world wide web. International World Wide Web Conferences Steering Committee, 2016: 107-108. DOI: 10.1145/2872518.2889368

45. Sumner C, Byers A, Boochever R, Park GJ. Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets. Machine learning and applications icmla, 2012 11th international conference on. Vol. 2. IEEE, 2012: 386-393. DOI: 10.1109/icmla.2012.218

46. Thilakaratne M, Weerasinghe R, Perera S. Knowledge-Driven Approach to Predict Personality Traits by Leveraging Social Media Data. Web Intelligence WI, 2016 IEEE/WIC/ACM International Conference on. IEEE, 2016: 288-295. DOI: 10.1145/2702123.2702280

47. Wald R, Khoshgoftaar T, Sumner C. Machine prediction of personality from Facebook profiles. Information Reuse and Integration IRI, 2012 IEEE 13th International Conference

on. IEEE, 2012: 109-115. DOI: 10.1109/iri.2012.6302998

48. Wei H, Zhang F, Yuan NJ, et al. Beyond the Words: Predicting User Personality from Heterogeneous Information. Proceedings of the Tenth ACM International Conference on Web Search and Data Mining. ACM, 2017: 305-314. DOI: 10.1145/3018661.3018717

49. Schwartz HA, Sap M, Kern ML, et al. Predicting individual well-being through the language of social media. Biocomputing 2016: Proceedings of the Pacific Symposium. 2016: 516 - 21. DOI: 10.1142/9789814749411_0047

50. Tsugawa S, Kikuchi Y, Kishino F, Nakajima K, Itoh Y, Ohsaki H. Recognizing depression from twitter activity. Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, 2015: 3187-3196. DOI: 10.1145/2702123.2702280

51. Wang X, Zhang C, Ji Y, Sun L, Wu L, Bao Z. A depression detection model based on sentiment analysis in micro-blog social network. Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer, Berlin, Heidelberg, 2013: 201-213. DOI: 10.1007/978-3-642-40319-4_18

52. He Q, Glas CA, Kosinski M, Stillwell DJ, and Veldkamp B.P. Predicting self-monitoring skills using textual posts on Facebook. Computers in human behavior 33, 2014: 69-78. DOI: 10.1016/j.chb.2013.12.026

53. Chen J, Hsieh G, Mahmud JU, and Nichols J. Understanding individuals' personal values from social media word use. Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing. ACM, 2014: 405-414.

54. Golbeck J. Detecting Coping Style from Twitter. International Conference on Social Informatics. Springer International Publishing, 2016: 454-467. DOI: 10.1007/978-3-319-

47880-7_28

55. Rosenthal R. The file drawer problem and tolerance for null results. Psychological bulletin 86.3, 1979: 638.

56. Confessore N, Hakim D. Data Firm Says 'Secret Sauce' Aided Trump: Many Scoff. New York Times, 6, March 2017, www.nytimes.com/2017/03/06/us/politics/cambridge-analytica.html?_r=0. Accessed on 25th June, 2017.

57. Cadwalladr C. The great British Brexit robbery: How our democracy was hijacked. The Guardian, May 7, 2017, www.theguardian.com/technology/2017/may/07/the-great-british-brexit-robbery-hijacked-democracy. Accessed 2nd July 2017.

58. Levin S. Facebook told advertisers it can identify teens feeling 'insecure' and 'worthless'. The Guardian, May 1, 2017, www.theguardian.com/technology/2017/may/01/facebook-advertising-data-insecure-teens. Accessed 26th June 2017.

59. Meyer, GJ, et al. Psychological testing and psychological assessment: A review of evidence and issues. American psychologist 56.2, 2001: 128.

60. Roberts BW, Kuncel NR, Shiner R, Caspi A, Goldberg LR. The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. Perspectives on Psychological Science 2.4, 2007: 313-345.

61. Kosinski M, Wang Y, Lakkaraiu H, Leskovec J. Mining big data to extract patterns and predict real-life outcomes. Psychological methods 21.4, 2016 493.

62. Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS. Deep learning for visual understanding: A review. Neurocomputing 187, 2016: 27-48.

63. Statista. The Most popular mobile social networking apps in the United States. May 2017, www.statista.com/statistics/248074/most-popular-us-social-networking-apps-ranked-by-audience/. Accessed 2nd July, 2017.

Table 1. Characteristics of Studies Included in the Systematic review and Meta-Analyses

Study (Effect)	Main Characteristic	Specific Characteristic	Self-report	r	N	Source (Quality)	Social media	Digital traces
Bachrach et al., 2012 *	Personality	Big 5 Traits	International Personality Item Pool Questionnaire (IPIP)	0.40	5000	Proceeding (Low)	Facebook	Activity
Bai et al., 2014 1	Psychological well-being	Life Satisfaction	Urban and Rural Residents Social Attitudes Questionnaire	0.30	2018	Repository (Low)	Sina Weibo	Demographics, Activity, Language
Bai et al., 2014 2	Social satisfaction	Income, Social Position, National Economy, Local Economy, Social Justice, Average Satisfaction.	(Same as above)	0.48	2018	Repository (Low)	Sina Weibo	Demographics, Activity, Language
Celli et al., 2014 1	Personality	Big 5 Traits	Big 5 Personality Test (BFI-10)	0.15	89	Proceeding (Low)	Facebook	Pictures
Celli et al., 2014 2	Personal values	Dominance & Affect	Interpersonal Circumplex (IPIP-IPC-32)	0.16	89	Proceeding (Low)	Facebook	Pictures
Chen et al., 2014	Personal values	Self-Transcendence, Self-Enhancement, Conservation, Openness to Change, Hedonism	Portrait Value Questionnaire (PVQ)	0.39	799	Proceeding (High)	Reddit	Language
De Choudury et al., 2013	Psychological well-being	Depression	CES-D & Beck Depression Inventory	0.48	476	Proceeding (Low)	Twitter	Demographics, Activity, Language
Farnadi et al., 2016 1 *	Personality	Big 5 Traits	IPIP	0.22	3731	Journal (High)	Facebook	Demographics, Activity, Language
Farnadi et al., 2016 3 *	Personality	Big 5 Traits	Big 5 Inventory - 10	0.37	44	Journal (High)	Twitter	Demographics
Gao et al., 2013	Personality	Big 5 Traits	44-Item Big 5 Personality Inventory	0.36	176	Proceeding (Low)	Sina Weibo	Activity, Language
Garcia & Sikstrom, 2014	Personality	Dark Triad, Extraversion & Neuroticism	Eysenck Personality Questionnaire Revised, Narcissistic Personality Inventory & Mach-IV	0.14	304	Journal (High)	Facebook	Language
Golbeck , 2016 b	Copying Style	Coping Style	Ways of Coping Survey	0.59	105	Proceeding (Low)	Twitter	Language
Golbeck et al., 2011	Personality	Big 5 Traits	45-Item Big 5 Personality Inventory	0.57	167	Proceeding (Low)	Facebook	Demographics, Activity, Language
Golbeck, 2016 1 *	Personality	Big 5 Traits	100-Item IPIP	0.35	127	Proceeding (Low)	Facebook	Language
Golbeck, 2016 2 *	Personality	Big 5 Traits	100-Item IPIP	0.21	8569	Proceeding (Low)	Facebook	Language
Golbeck, 2016 3 *	Personality	Big 5 Traits	45-Item Big 5 Personality Inventory	0.24	69	Proceeding (Low)	Facebook	Language

Gosling et al., 2011	Personality	Big 5 Traits	TIPI	0.25	133	Journal (High)	Facebook	Activity
He et al., 2014 *	Self-Monitoring	Self-monitoring Skills	Snyder's Self Monitoring Questionnaire	0.19	1128	Journal (High)	Facebook	Language
Kern et al., 2014 *	Personality	Big 5 Traits	IPIP	0.15	69792	Journal (High)	Facebook	Language
Kleanthous et al., 2016	Personality	Big 5 Traits	50-Item IPIP	0.15	62	Proceeding (Low)	Facebook	Activity
Kosinski et al., 2013 1 *	Personality	Big 5 Traits	IPIP	0.35	54373	Journal (High)	Facebook	Likes
Kosinski et al., 2013 2 *	Psychological well-being	Satisfaction with Life	Satisfaction with Life Scale	0.17	2340	Journal (High)	Facebook	Likes
Kosinski et al., 2013 3 *	Intelligence	Intelligence	Raven's Standard Progressive Matrices (SPM)	0.39	1350	Journal (High)	Facebook	Likes
Kosinski et al., 2013 4 *	Substance use	Smokes cigarettes, Drinks alcohol & Uses Drugs	Online Surveys – Not specified	0.34	856 – 1211	Journal (High)	Facebook	Likes
Kosinski et al., 2014 1 *	Personality	Big 5 Traits	IPIP	0.17	9515 – 45565	Journal (High)	Facebook	Activity
Kosinski et al., 2014 2 *	Psychological well-being	Satisfaction with Life	Satisfaction with Life Scale	0.33	311	Journal (High)	Facebook	Activity
Kosinski et al., 2014 3 *	Intelligence	Intelligence	Raven's SPM	0.2	395	Journal (High)	Facebook	Activity
Li et al., 2014	Personality	Big 5 Traits	Chinese Version of the 44-Item Big 5 Personality Inventory	0.54	547	Journal (High)	Sina Weibo	Activity
Liu et al., 2015 *	Psychological well-being	Satisfaction with Life	Satisfaction with Life Scale	0.15	1124	Journal (High)	Facebook	Language
Liu et al., 2016 1	Personality	Big 5 Traits	IPIP	0.19	254	Proceeding (Low)	Twitter	Language
Liu et al., 2016 2	Personality	Big 5 Traits	IPIP	0.12	429	Proceeding (Low)	Twitter	Pictures
Markovikj et al. 2013 *	Personality	Big 5 Traits	IPIP	0.67	250	Proceeding (Low)	Facebook	Demographics, Activity, Language
Park et al., 2015 *	Personality	Big 5 Traits	IPIP	0.38	4824	Journal (High)	Facebook	Language
Preotiuc-Pietro et al., 2016	Personality	Dark Triad	Dirty Dozen 12-Item Questionnaire	0.25	491	Proceeding (High)	Twitter	Language, Pictures
Qiu et al., 2012	Personality	Big 5 Traits	44-Item Big 5 Personality Inventory	0.22	142	Journal (High)	Twitter	Language
Quercia et al., 2012 1 *	Personality	Big 5 Traits	IPIP	0.08	2165	Proceeding (High)	Facebook	Activity
Quercia et al., 2012 2 *	Self-Monitoring	Self-Monitoring Skills	Snyder's Self-Monitoring Questionnaire	0.089	2165	Proceeding (High)	Facebook	Activity
Schwartz et al., 2013 *	Personality	Big 5 Traits	IPIP	0.35	18177	Journal (High)	Facebook	Language
Schwartz et al., 2014 *	Psychological well-being	Depression	Average response to 7 Depression Facet Items from the Neuroticism Item Pool (IPIP)	0.39	1000	Proceeding (Low)	Facebook	Language
Schwartz et al., 2016 *	Psychological well-being	Satisfaction with Life	Satisfaction with Life Scale and P.E.R.M.A Scale	0.30	440	Proceeding (Low)	Facebook	Language

Settanni & Marengo, 2015	Psychological well-being	Anxiety, Depression & Stress	Adapted Version of DASS-21	0.32	201	Journal (High)	Facebook	Language
Skowron et al., 2016	Personality	Big 5 Traits	44-Item Big 5 Personality Inventory	0.66	62	Proceeding (Low)	Twitter & Instagram	Language, Pictures
Sumner et al., 2012	Personality	Big 5 Traits & Dark Triad	TIPI & Short Dark Triad (SD3) Questionnaire	0.2	616	Proceeding (Low)	Twitter	Activity, Language
Thilakaratne et al., 2016 *	Personality	Big 5 Traits	IPIP	0.38	1000	Proceeding (Low)	Facebook	Language
Tsugawa et al., 2015 3	Psychological well-being	Depression	CES-D	0.32	209	Proceeding (High)	Twitter	Activity, Language
Wald et al., 2012	Personality	Big 5 Traits	45-Item Big 5 Personality Index	0.68	537	Proceeding (Low)	Facebook	Demographics, Activity, Language
Wang et al., 2013	Psychological well-being	Depression	Clinical Psychological Diagnosis	0.69	180	Proceeding (Low)	Sina Weibo	Activity, Language
Wei & Stillwell, 2016 *	Intelligence	Intelligence	Raven's SMP	0.27	7000	Repository (Low)	Facebook	Pictures
Wei et al., 2017	Personality	Big 5 Traits	44-Item Big Personality Inventory	0.40	949	Proceeding (Low)	Sina Weibo	Activity, Language, Pictures
Youyou et al., 2015 *	Personality	Big 5 Traits	100-Item IPIP	0.43	1919	Journal (High)	Facebook	Likes

Note: Studies included in the meta-analyses are in bold. * Study using MyPersonality datasets

Table 2
Independent Characteristics of Articles Included in the Meta-Analysis of Digital Traces, Social Media, and Psychosocial Characteristics: Study Characteristics

Study Characteristics	Frequency
Effect-sizes	(n = 50)
Personality	30
Big 5 traits	29
Dark-Triad	3
Psychological well-being	10
Depression	4
Emotional distress	1
Satisfaction with Life	5
Intelligence	3
Other psychosocial characteristics	6
Social satisfaction	1
Personal values	2
Copying style	1
Self-monitoring skills	2
Substance use	1
Social Media Platform	(n = 50)
Facebook	33
Twitter	10
Sina-Weibo	6
Instragram	1
Reddit	1
Type of Digital Traces	(n = 50)
Language	31
Activity Statistics	21
Likes	7
Demographics	8
Pictures	5
Analytic Approach	(n = 50)
Multiple Features	14
Single features	36
My Personality Studies	26
Sample Size	(n = 4
Less than 200	10
201-500	12
501-1,000	6
1,001-5,000	12
5,001-10,000	3
More than 10,000	6
Publication Source	(n = 38)

Online repository	2
Proceedings	16
Journal	20

Table 3: Results of univariate meta-regressions: factors moderating effect-size for personality (k=18) and psychological well-being (k=9)

	$\mathcal{T}^2$	R^2	β	S.E.	95% C.I.	P-Value
Personality						
Study quality (High vs. Low)	0.04	0.04	-0.08	0.10	-0.28 – 0.11	0.41
Private vs. Public settings	0.04	0.00	0.01	0.1	-0.19 – 0.21	0.95
Multiple vs. single types of digital traces	0.03	0.19	0.19	0.09	0.01 - 0.38	0.04
User demographics (Yes vs. No)	0.03	0.22	0.23	0.11	0.01 - 0.44	0.04
User activity statistics (Yes vs. No)	0.03	0.16	0.15	0.09	-0.04 – 0.33	0.12
Language (Yes vs. No)	0.04	0.03	0.1	0.11	-0.11 – 0.30	0.37
Pictures (Yes vs. No)	0.04	0.01	-0.02	0.11	-0.24 – 0.20	0.83
Psychological Well-Being						
Study quality (High vs. Low)	0.04	0.04	-0.08	0.10	-0.28 – 0.11	0.41
Private vs. Public settings	0.02	0.31	-0.18	0.10	-0.37– 0.01	0.06
Multiple vs. single types of digital traces	0.02	0.31	0.18	0.10	0.01 – 0.38	0.06
User activity statistics (Yes vs. No)	0.03	0.16	0.15	0.09	-0.04 – 0.33	0.12

List of figure legends

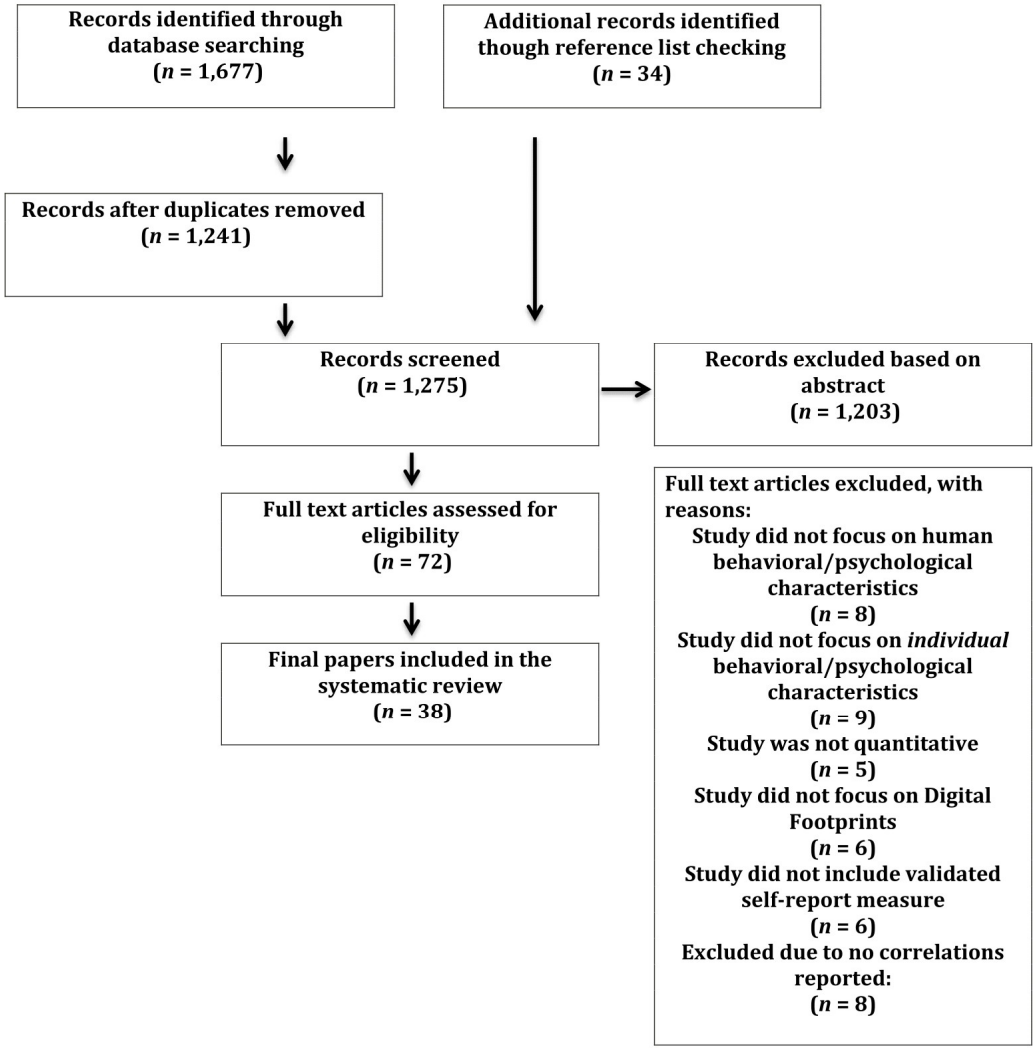

Figure 1: Flowchart of Article Selection.

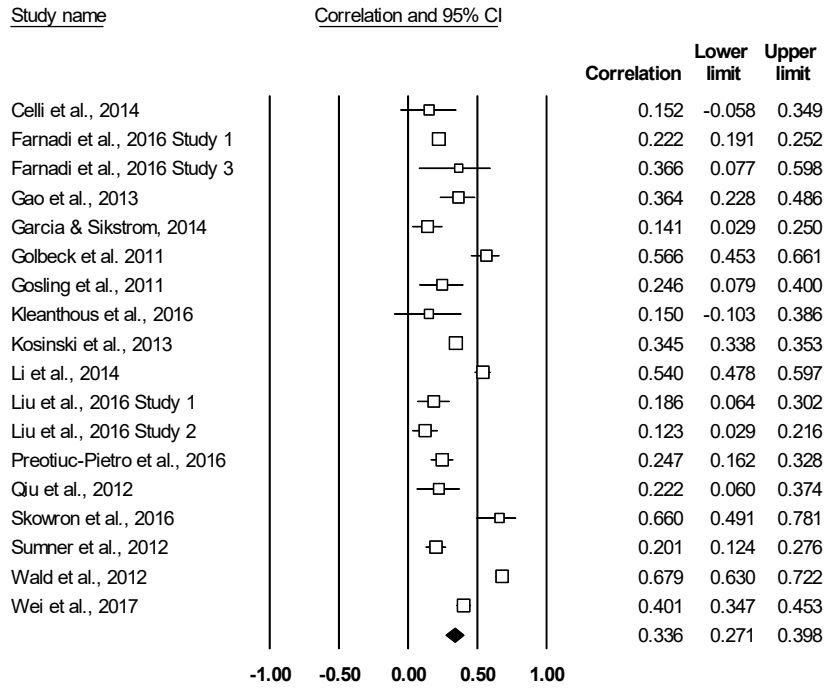

Figure 2 Forest plot of personality study-average effect sizes by weight

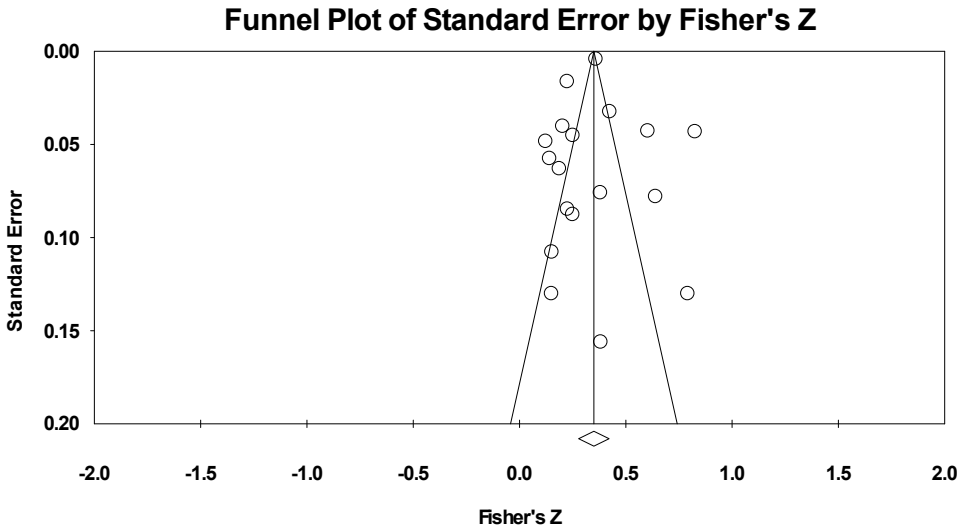

Figure. 3. Funnel plot displaying effect sizes for personality by SEs

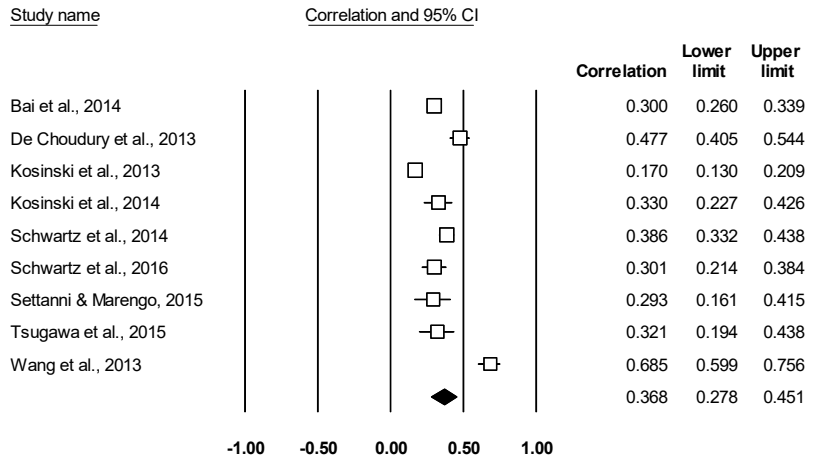

Figure 4 Forest plot of psychological well-being study-average effect sizes by weight

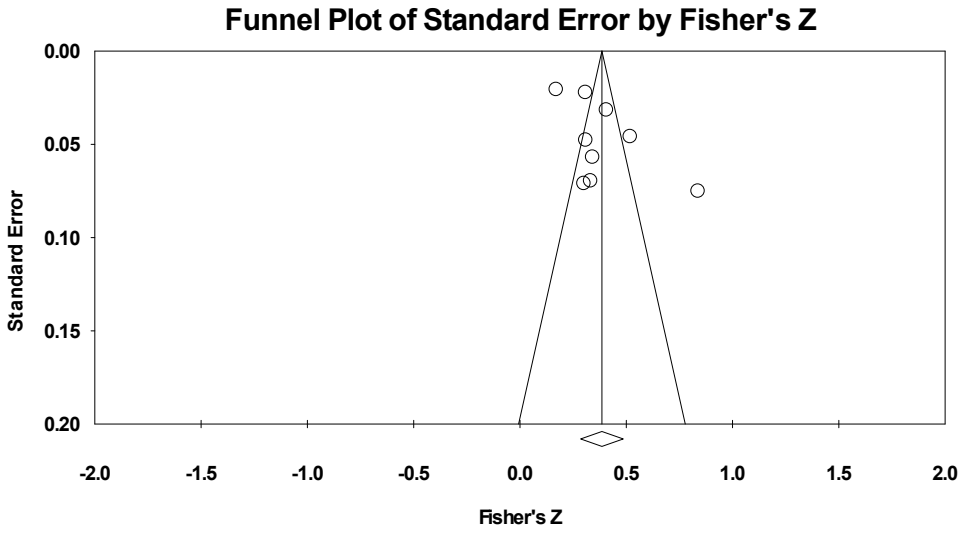

Figure. 5. Funnel plot displaying effect sizes for psychological well-being by SEs

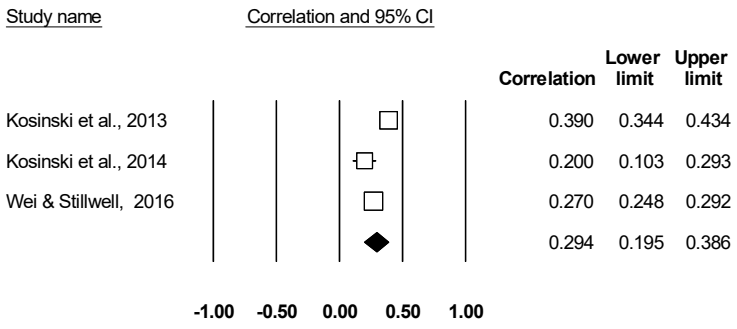

Figure 6 Forest plot of intelligence study-average effect sizes by weight

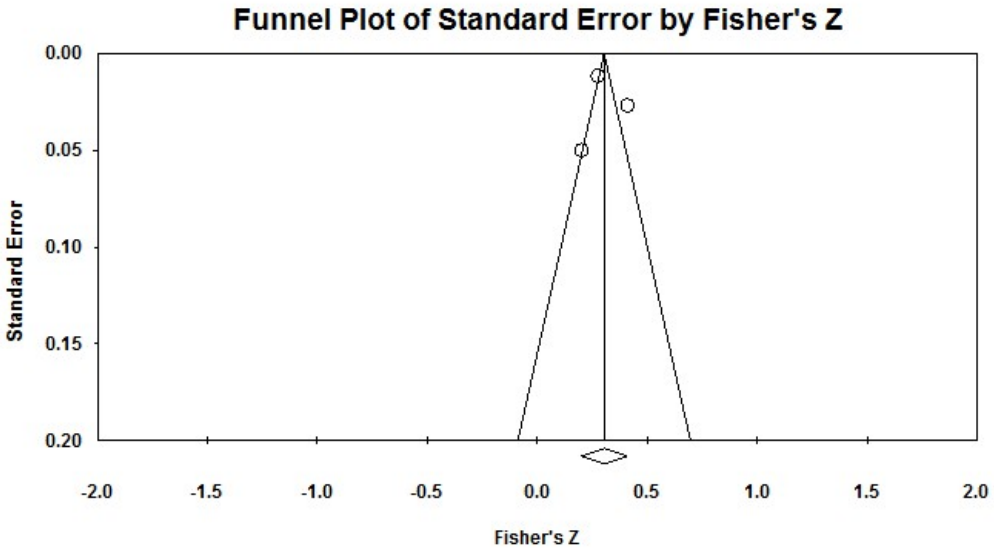

Figure. 7. Funnel plot displaying effect sizes for intelligence by SEs