

Can Psychological Traits Be Inferred From Spending? Evidence From Transaction Data

Psychological Science
 2019, Vol. 30(7) 1087–1096
 © The Author(s) 2019
 Article reuse guidelines:
sagepub.com/journals-permissions
 DOI: 10.1177/0956797619849435
www.psychologicalscience.org/PS



Joe J. Gladstone¹ , Sandra C. Matz², and Alain Lemaire²

¹UCL School of Management, University College London, and ²Columbia Business School, Columbia University

Abstract

The automatic assessment of psychological traits from digital footprints allows researchers to study psychological traits at unprecedented scale and in settings of high ecological validity. In this research, we investigated whether spending records—a ubiquitous and universal form of digital footprint—can be used to infer psychological traits. We applied an ensemble machine-learning technique (*random-forest modeling*) to a data set combining two million spending records from bank accounts with survey responses from the account holders ($N = 2,193$). Our predictive accuracies were modest for the Big Five personality traits ($r = .15$, corrected $\rho = .21$) but provided higher precision for specific traits, including materialism ($r = .33$, corrected $\rho = .42$). We compared the predictive accuracy of these models with the predictive accuracy of alternative digital behaviors used in past research, including those observed on social media platforms, and we found that the predictive accuracies were relatively stable across socioeconomic groups and over time.

Keywords

Big Five personality traits, consumer psychology, psychometrics, computational social science, financial decision making

Received 5/25/18; Revision accepted 3/17/19

The automatic prediction of psychological traits from digital footprints offers the potential to transform the scientific investigation of individual differences by allowing researchers to study psychological traits at unprecedented scale and in settings of high ecological validity. Driven by advances in computational methods and the wider availability of user-generated data (Kosinski, Matz, Gosling, Popov, & Stillwell, 2015), this research suggests that psychological traits can be inferred from digital records of behavior, including Facebook profiles (Kosinski, Stillwell, & Graepel, 2013; Park et al., 2014; Youyou, Kosinski, & Stillwell, 2014), Twitter profiles (Golbeck, Robles, Edmondson, & Turner, 2011), Flickr pictures (Segalin, Perina, Cristani, & Vinciarelli, 2017), and even music collections (Nave et al., 2018).

We extended this research by investigating whether patterns in a person's spending can also reveal differences in psychological traits. After all, spending is often considered a reflection of who people are as individuals (self-congruity theory; Sirgy, 1985): Consumers buy

products not only for what those products can do, but also for what they mean to them (Levy, 1959). With more than 20 billion payment cards in circulation (The Nilson Report, 2018), aggregated spending records provide a detailed metric of people's tastes and preferences—and this, combined with evidence that categories of spending have consistent associations with personality (Matz, Gladstone, & Stillwell, 2016), suggests that it may be possible to accurately infer a person's psychological profile using his or her spending records.

However, it is unclear how the accuracy of spending-records analysis will compare with the accuracy of analyses of other types of digital footprints. In fact, there are competing hypotheses that can be derived

Corresponding Author:

Joe J. Gladstone, UCL School of Management, University College London, Level 38, 1 Canada Square, E14 5AA, London, United Kingdom
 E-mail: j.gladstone@ucl.ac.uk

from the theoretical distinction between two types of observable behavioral traces, both of which have been shown to hold valid cues to a person's psychological profile: *identity claims* and *behavioral residues* (Gosling, Ko, Mannarelli, & Morris, 2002). Social media platforms are designed as a way for individuals to explicitly communicate their identity and preferences to others within their social network. Consequently, digital traces observed on social media can be thought of as identity claims: deliberate symbolic statements made by individuals to express themselves (Gosling et al., 2002). Digital music libraries or spending records, on the other hand, can be better described as behavioral residues: subtle cues about people's preferences inadvertently conveyed as a result of their activity (Gosling et al., 2002). On the one hand, because spending is recorded passively and includes information often hidden from other people, it might be less influenced by social desirability and might therefore be a more accurate reflection of people's psychological traits. On the other hand, spending records, like other behavioral residues, may be less predictive of people's psychological traits because they are weaker indicators of how people perceive themselves and desire to be perceived by others.

We explored this question by comparing the predictive accuracy of models built using spending records with the predictive accuracy of models built using alternative digital behaviors, as used in past research. We also extended past research by investigating whether predictive accuracies from models built from spending records are biased against certain groups, such as people who are poorer (and therefore spend less).

Method

To investigate whether personality can be predicted from spending, we combined two million spending records from bank accounts with survey responses completed by the account holders ($N = 2,193$). Using a predictive machine-learning approach (Yarkoni & Westfall, 2017), we examined the accuracy of out-of-sample predictions of participants' personalities and provided comparative estimates for how effectively different traits can be predicted from spending records. All customer data were fully anonymized, and we received ethical approval for the project from the university ethics committee.

Participants

The data set was collected in May 2017 in collaboration with a UK-based money management app. The service provides users with a single dashboard of their finances by aggregating transactions from all their bank accounts

and providers. For example, if a participant had two credit cards and one checking account, each with different service providers, then data from each of these accounts were recorded by the application. This pooling of account information represents an advantage over previous methodology used in bank-data research, which has typically relied on information derived from only a single bank (e.g., Matz et al., 2016).

Customers of the service were sent a survey link by e-mail asking them to take part in the study, with the opportunity to win a tablet computer as a prize. In the survey, participants consented to match their survey responses with their transaction data for research purposes. In total, 2,193 people completed the study and provided consent to participate. For 1,875 of those participants, the service provided information on age ($M = 38.07$ years, $SD = 11.46$) and annual income. Gender was not measured directly but was derived by running first names of account users through a names database, providing gender in just over half of cases (11% female, 43% male, 46% unknown).

Survey measures

Past research has focused primarily on using the Big Five personality traits (Costa & McCrae, 1992), but in this project, we chose to predict two additional traits that researchers have found to be more closely related to consumption: materialism and self-control. People with materialistic tendencies prefer material goods over experiential ones (Howell, Pchelin, & Iyer, 2012), and people with greater self-control spend less on impulsive purchases and save more (Oaten & Cheng, 2007). Thus, we expected spending records to provide more accurate predictions for these more specific and relevant traits than for the broader Big Five dimensions.

Big Five personality traits. The most widely accepted model of personality, the Big Five model (Costa & McCrae, 1992), proposes a taxonomy of five personality traits: Openness (open to new experience and complex vs. conventional and uncreative), Conscientiousness (dependable and self-disciplined vs. disorganized and careless), Extraversion (extraverted and enthusiastic vs. reserved and quiet), Agreeableness (sympathetic and warm vs. critical and quarrelsome), and Neuroticism (anxious and easily upset vs. calm and emotionally stable). To measure each trait, we used the 10-item version of the Big Five Inventory (BFI-10), an established short scale of the Big Five model (Rammstedt & John, 2007). The correlations (r s) between trait item pairs ranged from .20 for Agreeableness to .59 for Extraversion. With Cronbach's alphas ranging from .33 for Agreeableness to .75 for Extraversion, the internal consistencies of scales ranged from poor to acceptable. Given

the relatively low internal consistencies of our outcome measures, we accounted for measurement error in our analyses by correcting for attenuation.

Materialism. *Materialism* refers to the degree to which an individual considers material possessions and physical comfort important, and it was measured using three items taken from a widely used measure of materialism (Richins & Dawson, 1992): (a) “I admire people who own expensive homes, cars and clothes”; (b) “I like a lot of luxury in my life”; and (c) “I’d be happier if I could afford to buy more things.” Scale reliability was found to be low (Cronbach’s $\alpha = .62$). As with the Big Five traits, we corrected for attenuation in our analysis.

Self-control. *Self-control* refers to the degree to which individuals can control and regulate their impulses. The construct was measured using a single item (“I am good at resisting temptation”) from the Brief Self-Control Scale (Tangney, Baumeister, & Boone, 2004). The nature of our data collection required us to keep the number of questions to a minimum, but we believe that the single-item measure, although not ideal, provides a pragmatic approximation of the wider construct of self-control.

Spending records

The transaction records obtained from participants’ banking data encompassed detailed information of all purchases made using customers’ bank accounts (checking accounts and credit cards) over 12 months.

Broad spending categories. Individuals’ purchases were automatically grouped by the company into 279 categories, including supermarkets, furniture stores, and insurance policies. The full list of spending categories with the average annual amount spent and number of purchases per category is provided in Table S1 in the Supplemental Material available online. For the purposes of our analysis, we aggregated participants’ spending in each of the categories over the year preceding the survey.

We used participants’ relative spending across categories, rather than the raw amounts, to help ensure that our predictors reflected patterns in spending rather than each participant’s total income or wealth. To calculate this, we divided participants’ spending in an individual category by their overall spending, giving us the relative amount spent in each specific category.¹ We then centered each of the 279 features before using them in the prediction models. It is worth noting that because we expected wealthier individuals to spend a lower proportion of their overall budget on necessities, our proportional measure of spending might still vary across the resource levels of the participants. (We tested

the implications of this idea on the accuracy of our models more explicitly in the context of moderation analyses.)

Specific merchant names. Customer purchases were tagged with the name of one of 658 identifiable merchants (e.g., Tesco or Amazon); the transaction was assigned a “no-merchant” tag if the merchant could not be identified. The merchants that were represented most frequently in our data set were Tesco (UK supermarket; 5.68%), Amazon (online retailer; 4.01%), Sainsbury (UK supermarket; 3.80%), and PayPal (money-transfer service; 3.36%).

To reduce the dimensionality of the merchant-tag data, we applied latent Dirichlet allocation (LDA), an established data-clustering technique within probabilistic machine learning. While LDA is most commonly used in classifying text, it can also be used in contexts such as ours, in which the goal was to group data on the basis of shared characteristics. LDA assumes that certain merchant tags frequently appear together and form a coherent theme or topic. In our data, for example, the merchants McDonald’s, KFC, Pizza Hut, Subway, and Burger King form a single topic (which we labeled “fast food”), whereas Costa, Starbucks, and Caffè Nero form a separate topic (which we labeled “coffee shops”). LDA then assumes that the spending habits of each person can be described as a weighted combination of these themes. For example, one person’s transactions could be a combination of 80% fast food and 20% coffee, while another person’s spending could be made up of 50% each.

To apply LDA to our data, we aggregated all the identifiable merchant tags used by each customer during our observation period into a “bag of tags.”² Following the recommended procedure for identifying the optimal number of topics, we split the bags of tags into a standard 80-20 training-test set and used the perplexity score—an out-of-sample measure of model fit—as the criterion to evaluate each model. The analyses yielded a total of 34 topics in our data set. In addition to the fast-food and coffee-shop topics, we also identified topics related to investment services, utilities, and electronics (an interactive chart of topics is available at https://osf.io/x3gwe/?view_only=ebbe995db96947ebb33c1a67e738b2bc). These 34 topics, as well as the 279 broad spending categories, were included as features in our predictive model.

Prediction models

Machine-learning algorithms provide new opportunities for researchers in psychology to gain valuable insights from large-scale behavioral data. The traditional psychological toolbox has provided tools well calibrated

to analyze the results from classic experimental paradigms (e.g., when comparing a limited number of experimental conditions), but it has yet to provide appropriate tools for analyzing large-scale observational data sets of human behavior (i.e., *big data*). To predict psychological traits from participants' transaction histories, we employed *random-forest models*. These models are widely used in other disciplines, including computer-science research, but remain rare in psychological research. Thus, to make the method accessible to our readers—and to encourage other researchers to use these methods in their own work—we first provide a brief explanation of random-forest models and then describe how we applied the model to our data.

An introduction to random-forest models. To grasp the logic behind random-forest models, it is necessary to start with an explanation of their essential building blocks: decision trees. Although we rarely label them as such, decision trees are simply a way of describing the logic people use to make decisions on a daily basis. Take a clinical psychologist, for example, who is trying to diagnose a patient suffering from a list of symptoms. To recommend appropriate treatments, the psychologist needs to determine whether the patient is depressed and, if so, how severely. To classify the patient as mildly depressed, severely depressed, or not depressed at all, the psychologist asks questions such as, "Have you lost interest in activities which you used to enjoy?" These questions form the branches of the tree, and as the patient answers each of these questions, the therapist gets closer to a diagnosis. If the patient confirms a loss of interest in enjoyable activities, then the likelihood that the patient is suffering from depression increases. The therapist might decide to follow up with a further question (e.g., "Have you had suicidal thoughts?") aimed at confirming the initial hypothesis. After going through several of these decision trees, the therapist makes the diagnosis: The patient is indeed depressed, and severely so.

Just like their real-world counterparts, decision trees in random-forest models are aimed at narrowing down the set of possible outcomes. When trying to predict a person's Extraversion from the available transaction data, for example, the model could ask whether that person spends a large amount of money on dining and drinking, a spending category found to be correlated with Extraversion (Matz et al., 2016). If the answer is yes, then the likelihood of the participant being more extraverted than the average person increases; if the answer is no, the likelihood decreases. While the psychologist in our example decides on questions a priori, by using the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric

Association, 2013), the "questions" our random-forest model asks are instead developed by the model on the basis of the data used to train it. In our case, the model learned to map input data (spending records) to output data (personality scores) by learning how to build a decision tree that yielded the highest accuracy in mapping one to the other. This process can be compared with learning the optimal questions to ask when diagnosing depression.

Random-forest models are a combination of large numbers of decision trees. To explain why we need to combine trees together into a forest, let us return to the therapist who has diagnosed a patient as severely depressed. Was this diagnosis the correct one? Even with highly standardized diagnostic tools, there remains ample room for subjectivity and error in this type of decision. Indeed, more than 60% of patients diagnosed with depression by a clinician did not meet the official criteria for the disorder on reevaluation (Mojtabai, 2013). Therefore, should we trust the opinion of this one therapist or get a second opinion? Better yet, should we source diagnoses from a few hundred, or a few thousand, therapists? In the latter case, some therapists will diagnose the patient as mentally healthy, whereas others will diagnose him or her as severely depressed. By relying on the diagnosis most commonly suggested across this pool of psychologists, we are more likely to end up with the correct diagnosis, compared with relying on any individual therapist chosen at random. This simple concept of "the wisdom of the crowds," in which groups of people pool their abilities to show collective intelligence, is what underlies the predictive accuracy of random-forest models: A large number of decision trees are aggregated to improve predictive accuracy.

Finally, what makes random forests random? Each decision tree in a random-forest model has access to only a random subset of input data (e.g., spending categories) and a random subset of participants. This artificial increase in variance allows the final model to produce a more robust mapping of input and output data. This is similar to our clinical psychologist being provided with only a random subset of questions to ask patients and being able to draw only on a random subset of prior experiences to make a diagnosis.

Model specifications. On the basis of the logic of random-forest models outlined above, we fitted a model to our data that had the following two features. First, each decision tree was constructed using a different bootstrapped subsample of the data. Second, the split of each node in a tree was determined by a random subset of predictors. Combined, these two features make random-forest models particularly robust to overfitting.

To further reduce the risk of overfitting, we followed a standard 10-fold cross-validation protocol. Cross-validation involves training and testing a model on different samples of data, allowing the researcher to quantify the out-of-sample prediction error. This means we first randomly split the data set into 10 samples and then trained the random-forest model using 90% of the data set (9 training samples) to predict the Big Five traits plus materialism and self-control. This training sample was then separated into a 66% fitting sample and a 33% validation sample. Using the fitting and validation set, we performed an exhaustive grid search over the parameters of the random forest to determine the optimal model specification. During the grid-search process, we varied the following three parameters: (a) The number of trees was varied from 50 to 350 in increments of 50; (b) the maximum depth of the trees was varied between 5, 15, 30, 60, and 120; and (c) the number of predictors to consider at each split was varied from the $\log(k)$, \sqrt{k} , and k , where k is the number of predictors. In a second step, we used the best-trained model to predict the scores of participants in the remaining 10% of the data set (the holdout testing sample). In a third step, we estimated the predictive accuracy of our model by calculating Pearson product-moment correlations between the predicted and actual scores for the Big Five traits plus materialism and self-control. This three-step procedure was repeated 10 times, each time with different data in the training and testing sets.

Results

Three central questions motivated our research: (a) Can spending records predict a person's psychological traits? (b) do sociodemographic variables moderate the predictive accuracy of spending records? and (c) how does the predictive accuracy of spending records compare with that of other digital footprints?

Can spending records predict psychological traits?

Across all psychological traits measured in our study, the average correlation between actual and predicted scores was .19. However, this aggregated measure of accuracy masks considerable variation across individual traits, with Openness having the lowest accuracy ($r = .12$) and materialism having the highest ($r = .33$). The differences in predictive accuracy across the different traits suggests that transaction records provide greater predictive accuracy for the more focused psychological traits (materialism and self-control) than for the more general psychological traits (the Big Five). While the

average accuracy (r) for the narrow traits was .30, the average accuracy for the broad personality traits was only .15. If we correct for attenuation, a procedure that accounts for measurement error in the outcome variable, the average correlation (ρ) for all traits increases to .24. (Table S2 in the Supplemental Material lists the individual correlations, both with and without corrections for attenuation.)

To develop a more intuitive understanding of which spending categories were driving the predictive accuracy in each of the models, we calculated the univariate correlations between each of the spending categories and each of the psychological traits. Table S3 in the Supplemental Material lists the five categories most positively and negatively correlated with each trait. For example, Openness was found to be positively related to spending money on flights, Extraversion to dining and drinking, Agreeableness to donations, Conscientiousness to savings, and materialism to jewelry. Similarly, there was a negative association between self-control and bank charges, materialism and donations, and Neuroticism and mortgage payments. The direction of these correlations provides some face validity for the expected relationships between categories and personality, supporting prior associations found in consumer psychology: Conscientious individuals allocate more money to savings and investments, open-minded individuals spend more money on travel, and materialistic individuals give less to charity (Belk, 1985; Matz et al., 2016; Mosca & McCrory, 2016).

Returning to the overall predictive accuracies, we further explored the degree to which the predictive accuracy of our models depends on the amount of data available about a participant. In other words, would our predictions remain relatively stable if we had collected only a single month of data about a person, or do the random-forest models require the full year of transaction data? To answer this, we recalculated our models by training data on 90% of participants and predicting the scores of the remaining 10% using 1 to 11 months of their data. This process was repeated 10 times, so the resulting accuracies represent 10-fold cross-validated averages. Overall, the analysis suggests that the accuracy of the models increases if additional months of data are included. For example, the average predictive accuracy across all traits after 1 month ($r = .15$) was lower than after 6 months ($r = .19$). However, the stability of these associations also shows that a degree of accuracy can be obtained with a relatively limited amount of data (see Fig. S1 in the Supplemental Material).

For the purposes of comparison, we also built a model using the demographic information we had available for each user as predictors: age and gender. This

provided an average correlation (r) of .11 across all traits ($\rho = .17$), illustrating that spending accounted for a greater proportion of the variance in personality compared with a person's age and gender combined. However, it should be noted that while demographics were less predictive of personality than spending records across all traits, this was not universally the case; they were more accurate than spending records when predicting the Big Five trait of Agreeableness. When combined in a single model, demographics and spending records predicted the psychological traits with an average accuracy (r) of .20 ($\rho = .27$), outperforming the predictions of either spending records or demographics alone. One can think of this approach as a way to norm our prediction models, just as self-report questionnaires often provide separate norms for different sociodemographic groups (e.g., age and gender). For example, spending more money on jewelry might be predictive of higher levels of materialism in women, but it might be predictive of higher Agreeableness in men (who are more likely to buy jewelry for their loved ones). Adding age and gender as features allows the random-forest model to use those characteristics in building the trees and hence to develop idiosyncratic models for each sociodemographic group. The predictive validity of these models is provided in Table S3 and Figure S2 in the Supplemental Material.

Do sociodemographic variables moderate the predictive accuracy of spending records?

An important question when it comes to the accuracy of predictive algorithms is whether they can predict outcomes with a similar degree of accuracy across different sociodemographic groups. For example, it is possible that an algorithm that predicts personality from people's spending is more accurate for those who spend more overall. After all, these people have a greater number of purchases to evaluate, and they may also have more discretion over the purchases they make. To explore this question, we tested whether the absolute error of our predictions—that is, the absolute difference between predicted and actual scores—depends on the sociodemographic variables of age, salary (log transformed), total spending (log transformed), and deprivation level (log transformed). Although age, salary, and total spending were provided directly by the app, we calculated the deprivation level of participants' local area by matching their postal areas with UK census data. Consequently, the deprivation variable approximates a participant's deprivation level by using the deprivation level of the residential neighborhood.

To calculate a single measure of prediction error, we averaged the prediction error across the seven predictions we made for each person (five personality traits plus materialism and self-control). The results of a linear regression analysis suggest that, overall, the predictive accuracy is relatively stable across the sociodemographic variables we investigated. With an adjusted R^2 of .004, our moderators explained less than 1% of the variance in prediction error. The only variable that became significant at an alpha level of .05 was deprivation ($b = 0.25$, $SE = 0.11$, $\beta = 0.058$, $p = .028$), with participants living in areas that are highly deprived being more difficult to predict (see Table S4 in the Supplemental Material for the full model output). To develop a better understanding of how deprivation relates to predictive accuracy, we ran additional regression analyses predicting the absolute difference for each trait from the deprivation index. The effect went in the same direction for all traits, but it reached significance only for Neuroticism ($b = 0.55$, $SE = 0.25$, $\beta = 0.072$, $p = .029$) and self-control ($b = 0.33$, $SE = 0.14$, $\beta = 0.061$, $p = .018$). We visualize these effects in Figure 1, using surface-level plots, which illustrate the relationship between actual and predicted scores on these two traits as a function of deprivation.

One possible explanation for the moderating effect of postcode-level deprivation is that people living in poorer neighborhoods are less likely to have a large discretionary-spending budget to allocate to products and services that align with their personality. Instead, they may need to use their resources for necessities (e.g., groceries). However, given that we did not find significant effects for either total spending or income at the individual level, the findings might instead suggest that deprived areas offer fewer opportunities to spend money in a way that reflects psychological preferences. However, it should be noted that these differences are both practically small and statistically weak, so future research should aim to replicate these findings to ensure they are robust before greater emphasis is given to interpreting their meaning.

How does the predictive accuracy of spending records compare with that of other digital footprints?

To illustrate the relative accuracy of these prediction models, and to contextualize them within the broader literature of automatic personality prediction, we compared our results with those reported in other recent studies (see Fig. 2 and Table 1) predicting the Big Five traits from several online digital footprints, including Facebook Likes (Youyou et al., 2014), Facebook status updates (Park et al., 2014), Flickr pictures (Segalin et al.,

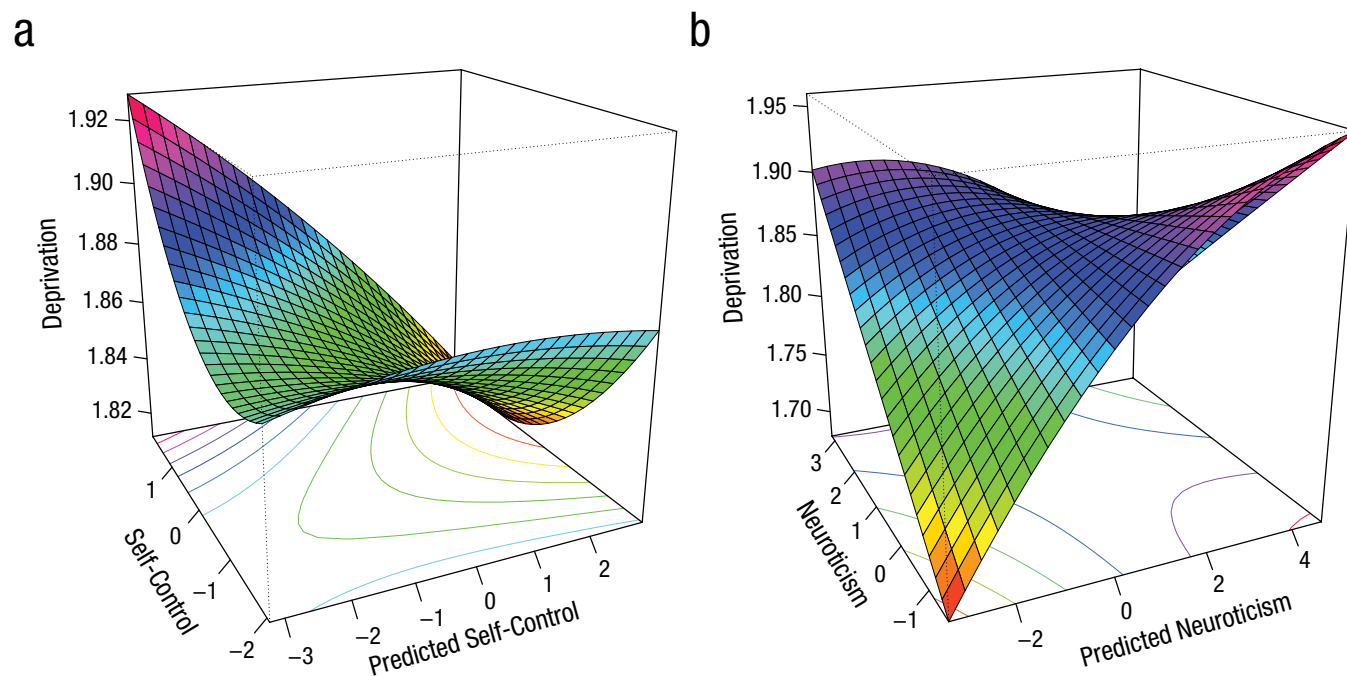


Fig. 1. Surface-level plots illustrating the relationship between actual and predicted scores for self-control (a) and Neuroticism (b), as moderated by deprivation.

2017), and music preferences (Nave et al., 2018). While the predictions from Facebook Likes and status updates outperformed our predictions considerably, our findings are comparable with those found in the context of music preferences and online photo-sharing websites.

These results are consistent with the distinction between identity claims and behavioral residues that we outlined in the introduction (Gosling et al., 2002). Facebook is designed for users to communicate their preferences and express themselves to others and therefore predominantly captures identity claims. In contrast, digital music libraries or spending records are far less curated and hence constitute more subtle behavioral residues. Although these comparisons of accuracies across studies are far from conclusive (as Fig. 2 and Table 1 illustrate, the studies differ on a number of dimensions), they suggest that behavioral residues might be less predictive of people's self-reported personalities than identity claims are. Spending records, specifically, provide a weaker social signal than other digital sources, but the lower accuracy might also be driven by the fact that spending patterns (unlike social media profiles) are not necessarily an expression of an individual's personal preferences: A large proportion of spending goes to fixed costs (e.g., groceries, bills) as well as to spending on others (e.g., partners, children). This may add additional noise to the transaction records, lowering the predictive accuracy of the

spending features in comparison with other types of behavioral footprints.

Discussion

Our findings contribute to research on the automatic prediction of psychological traits by illustrating that digital records of spending can be used to predict personality at unprecedented scale. Although our predictions include considerable error, they provide greater predictive accuracy for more focused psychological traits (materialism and self-control) and were generally robust across participants with different levels of financial resources, varying slightly on the basis of the deprivation levels of participant's local area on some traits.

Consistent with the theoretical distinction between behavioral residues and identity claims, predictive accuracies from spending records were modest compared with those found in past research using social media data (e.g., Youyou et al., 2014). However, given that these differences may also be due to sample composition, size, measurement error in the survey, the cultural characteristics of respondents, or other variables, there is a need for further research to differentiate between these behavioral footprints by collecting multiple digital records (e.g., Facebook Likes and spending records) from the same sample of participants.

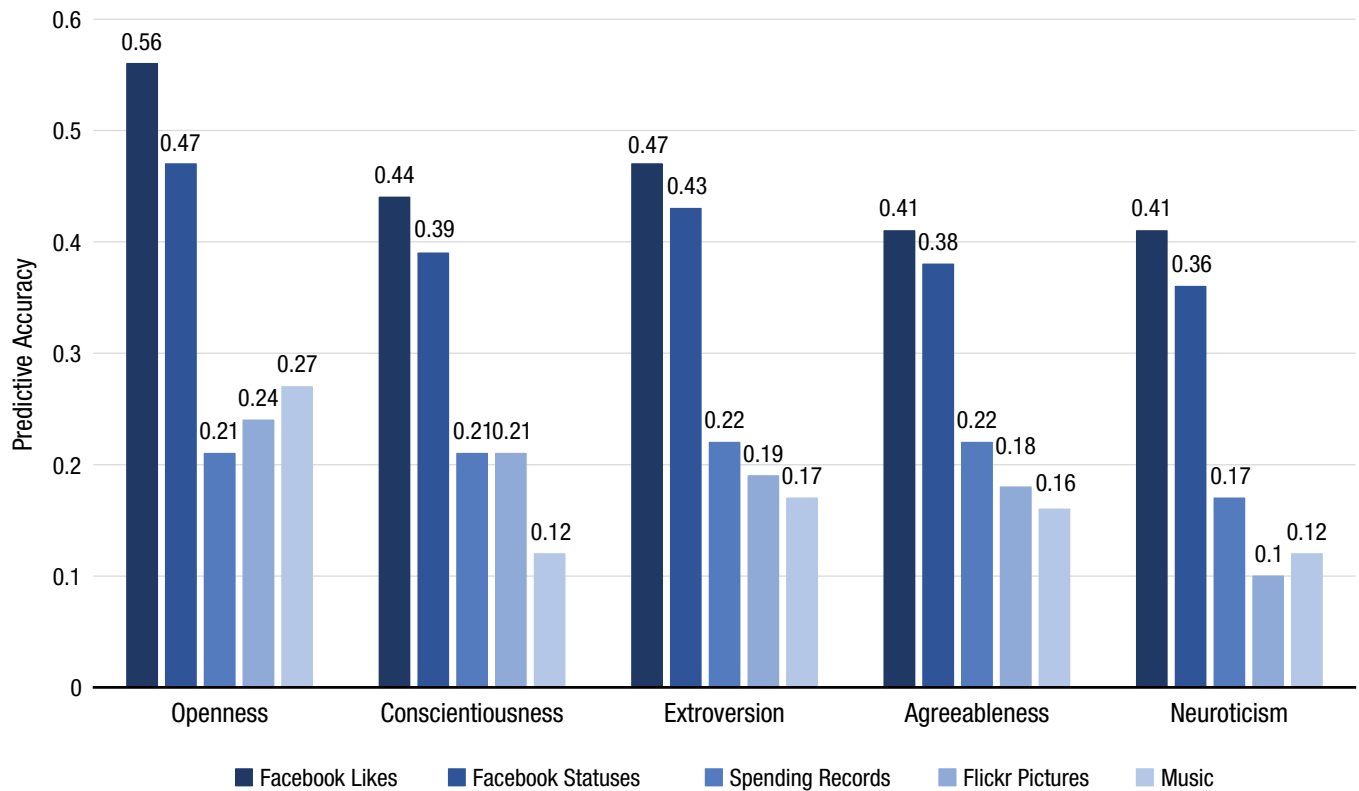


Fig. 2. Predictive accuracies of the Big Five traits, comparing Facebook Likes, Facebook statuses, spending records, Instagram pictures, and music preferences. Bars represent correlations accounting for attenuation, which we calculated using the alphas reported in the following studies: Facebook Likes (Youyou, Kosinski, & Stillwell, 2014), Facebook status updates (Park et al., 2014), spending records (this study), Flickr pictures (Segalin, Perina, Cristani, & Vinciarelli, 2017), and music preferences (Nave et al., 2018). Alphas were either reported in the original manuscripts or provided by the authors.

Predicting personality from spending also raises serious challenges for ethics and privacy. Firms could use personality predictions to identify and target vulnerable individuals, such as those low in self-control, with persuasive advertising for products harmful to their welfare or harmful to society at large (e.g., advertisements for gambling or smoking). Because personality predictions generated through one domain, such as spending records, can be used to target those same individuals

in others, such as through direct mail, it is increasingly difficult for individuals to escape this new form of automatic psychological assessment and its downstream applications. This means that as personality predictions become more accurate and ubiquitous, and as behavior is recorded digitally at an increasing scale, there is an urgent need for policymakers to ensure that individuals (and societies) are protected against potential abuse of such technologies.

Table 1. Comparison of Models and Samples From the Present Study With Those From Other Recent Studies

Variable	Facebook Likes	Facebook statuses	Spending records	Flickr pictures	Music
Personality measure (no. of items)	IPIP (20–100)	IPIP (20–100)	BFI-10 (10)	BFI-10 (10)	IPIP (20–100)
Sample size	86,220	71,556	2,193	300	22,252
Model	Lasso	Ridge	Random forest	Lasso	Lasso
Model tuning	Yes	Yes	Yes	Yes	Yes
Dimensionality reduction	No	Yes	Partly	Yes	No

Note: Data were obtained from the following studies: Facebook Likes (Youyou, Kosinski, & Stillwell, 2014), Facebook status updates (Park et al., 2014), spending records (this study), Flickr pictures (Segalin, Perina, Cristani, & Vinciarelli, 2017), and music preferences (Nave et al., 2018). Lasso and ridge are commonly used machine-learning models for linear regression analyses (Tibshirani, 1996). Model tuning indicates whether the parameters of the lasso and ridge regression models were optimized during the model-training stage. Dimensionality reduction indicates whether the original input matrix of digital footprints was collapsed to a smaller matrix using principal component analysis or singular value decomposition before training the models. BFI-10 = 10-item version of the Big Five Inventory (Rammstedt & John, 2007); IPIP = International Personality Item Pool.


Action Editor

Leaf Van Boven served as action editor for this article.

Author Contributions

J. J. Gladstone and S. C. Matz contributed equally to this study; author order is alphabetical. J. J. Gladstone and S. C. Matz developed the study concept. J. J. Gladstone collected the data and performed initial analyses. S. C. Matz and A. Lemaire performed substantial additional analyses. J. J. Gladstone and S. C. Matz drafted the manuscript. All authors approved the final version of the manuscript for submission.

ORCID iD

Joe J. Gladstone  <https://orcid.org/0000-0002-8783-3923>

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Supplemental Material

Additional supporting information can be found at <http://journals.sagepub.com/doi/suppl/10.1177/0956797619849435>

Open Practices

Data and materials for this study have not been made publicly available, and the design and analysis plans were not preregistered.

Notes

1. For example, if a participant had spent a total of £500 on books and £30,000 overall, the relative spending on books would be 1.7%, the same as for a person who spent only £50 on books and £3,000 overall.
2. If a participant purchased twice at Amazon.com and once at Apple, the individual's bag of tags would be "Amazon, Apple, Amazon."

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Belk, R. W. (1985). Materialism: Trait aspects of living in the material world. *Journal of Consumer Research*, *12*, 265–280. doi:10.1086/208515
- Costa, P., & McCrae, R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment*, *4*, 5–13. doi:10.1037//1040-3590.4.1.5
- Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011). Predicting personality from Twitter. In *Proceedings: 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing (PASSAT/SocialCom 2011)*; pp. 149–156). Washington, DC: IEEE. doi:10.1109/PASSAT/SocialCom.2011.33
- Gosling, S. D., Ko, S., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology*, *82*, 379–398. doi:10.1037//0022-3514.82.3.379
- Howell, R. T., Pchelin, P., & Iyer, R. (2012). The preference for experiences over possessions: Measurement and construct validation of the Experiential Buying Tendency Scale. *Journal of Positive Psychology*, *7*, 57–71. doi:10.1080/17439760.2011.626791
- Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist*, *70*, 543–556. doi:10.1037/a0039210
- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences, USA*, *110*, 5802–5805. doi:10.1073/pnas.1218772110.1218772110
- Levy, S. J. (1959). Symbols for sale. *Harvard Business Review*, *37*(4), 117–124.
- Matz, S. C., Gladstone, J. J., & Stillwell, D. (2016). Money buys happiness when spending fits our personality. *Psychological Science*, *27*, 715–725. doi:10.1177/09567976166635200
- Mojtabai, R. (2013). Clinician-identified depression in community settings: Concordance with structured-interview diagnoses. *Psychotherapy and Psychosomatics*, *82*, 161–169. doi:10.1159/000345968
- Mosca, I., & McCrory, C. (2016). Personality and wealth accumulation among older couples: Do dispositional characteristics pay dividends? *Journal of Economic Psychology*, *56*, 1–19. doi:10.1016/j.joep.2016.04.006
- Nave, G., Minxha, J., Greenberg, D. M., Kosinski, M., Stillwell, D., & Rentfrow, J. (2018). Musical preferences predict personality: Evidence from active listening and Facebook Likes. *Psychological Science*, *29*, 1145–1158. doi:10.1177/0956797618761659
- The Nilson Report. (2018). Issue 1140. Retrieved from https://nilsonreport.com/upload/issues/1140_0321.pdf
- Oaten, M., & Cheng, K. (2007). Improvements in self-control from financial monitoring. *Journal of Economic Psychology*, *28*, 487–501. doi:10.1016/j.joep.2006.11.003
- Park, G., Schwartz, H. A., Eichstaedt, J., Kern, M. L., Kosinski, M., Stillwell, D., . . . Seligman, M. E. P. (2014). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, *108*, 934–952.
- Rammstedt, B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, *41*, 203–212. doi:10.1016/j.jrp.2006.02.001
- Richins, M. L., & Dawson, S. (1992). A consumer values orientation for materialism and its measurement: Scale development and validation. *Journal of Consumer Research*, *19*, 303–316. doi:10.1086/209304
- Segalin, C., Perina, A., Cristani, M., & Vinciarelli, A. (2017). The pictures we like are our image: Continuous mapping of favorite pictures into self-assessed and attributed personality traits. *IEEE Transactions on Affective Computing*, *8*, 268–285. doi:10.1109/TAFFC.2016.2516994
- Sirgy, M. J. (1985). Using self-congruity and ideal congruity to predict purchase motivation. *Journal of Business*

- Research*, 13, 195–206. doi:10.1016/0148-2963(85)90026-8
- Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, 72, 271–324. doi:10.1111/j.0022-3506.2004.00263.x
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58, 267–288.
- Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, 12, 1100–1122. doi:10.1177/1745691617693393
- Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences, USA*, 112, 1036–1040. doi:10.1073/.1418680112.1418680112