Predicting personality from patterns of behavior collected with smartphones

Clemens Stachl[a,1], Quay Au[b], Ramona Schoedel[c], Samuel D. Gosling[d,e], Gabriella M. Harari[a], Daniel Buschek[f], Sarah Theres Völkel[g], Tobias Schuwerk[h], Michelle Oldemeier[c], Theresa Ullmann[i], Heinrich Hussmann[g], Bernd Bischl[b], and Markus Bühner[c]

[a]Department of Communication, Media and Personality Laboratory, Stanford University, Stanford, CA 94305; [b]Department of Statistics, Computational Statistics, Ludwig-Maximilians-Universität München, 80539 Munich, Germany; [c]Department of Psychology, Psychological Methods and Assessment, Ludwig-Maximilians-Universität München, 80802 Munich, Germany; [d]Department of Psychology, University of Texas at Austin, Austin, TX 78712; [e]School of Psychological Sciences, University of Melbourne, Parkville, VIC 3010, Australia; [f]Research Group Human Computer Interaction and Artificial Intelligence, Department of Computer Science, University of Bayreuth, 95447 Bayreuth, Germany; [g]Media Informatics Group, Ludwig-Maximilians-Universität München, 80337 Munich, Germany; [h]Department of Psychology, Developmental Psychology, Ludwig-Maximilians-Universität München, 80802 Munich, Germany; and [i]Institute for Medical Information Processing, Biometry, and Epidemiology, Ludwig-Maximilians-Universität München, 81377 Munich, Germany

Edited by Susan T. Fiske, Princeton University, Princeton, NJ, and approved June 18, 2020 (received for review December 2, 2019)

Smartphones enjoy high adoption rates around the globe. Rarely more than an arm's length away, these sensor-rich devices can easily be repurposed to collect rich and extensive records of their users' behaviors (e.g., location, communication, media consumption), posing serious threats to individual privacy. Here we examine the extent to which individuals' Big Five personality dimensions can be predicted on the basis of six different classes of behavioral information collected via sensor and log data harvested from smartphones. Taking a machine-learning approach, we predict personality at broad domain (r_{median} = 0.37) and narrow facet levels (r_{median} = 0.40) based on behavioral data collected from 624 volunteers over 30 consecutive days (25,347,089 logging events). Our cross-validated results reveal that specific patterns in behaviors in the domains of 1) communication and social behavior, 2) music consumption, 3) app usage, 4) mobility, 5) overall phone activity, and 6) day- and night-time activity are distinctively predictive of the Big Five personality traits. The accuracy of these predictions is similar to that found for predictions based on digital footprints from social media platforms and demonstrates the possibility of obtaining information about individuals' private traits from behavioral patterns passively collected from their smartphones. Overall, our results point to both the benefits (e.g., in research settings) and dangers (e.g., privacy implications, psychological targeting) presented by the widespread collection and modeling of behavioral data obtained from smartphones.

personality | behavior | machine learning | mobile sensing | privacy

It has been well documented that "digital footprints" derived from social network platforms (e.g., Facebook likes) can reveal individuals' psychological characteristics, such as their personality traits (1). This is consequential because the Big Five personality traits have been shown to predict a broad range of life outcomes in the domains of health, political participation, personal and romantic relationships, purchasing behaviors, and academic and job performance (2–4). Data-driven inferences about individuals' personality traits present great opportunities for research; but they also have major implications for individual privacy because they allow for personality-based targeting and manipulation (5, 6).

Even greater threats to privacy are posed by smartphones, which can collect a far broader, fine-grained array of daily behaviors than can be scraped from social media platforms and which are pervasive in most societies around the globe (7). The on-board sensors of a smartphone and the device's logging capabilities (e.g., app-usage logs, media and website consumption, location, communications, screen activity) can be harnessed by apps to record daily behaviors performed both on the devices themselves and in close proximity to them (8–10). These data have great potential for psychological research and have already begun to yield valuable findings, including studies relating physical activity and communication data to human emotion and mental wellbeing (11–14). However, behavioral data from smartphones can contain private information and should therefore be collected and processed only when informed consent is given (15). In theory, users must give permission for apps to access certain types of data on their phones (e.g., to record location or audio data). However, people are often unaware of the data they are providing, are tricked into giving access to more data (16), and struggle to understand current permission systems that are unspecific and ineffective in preventing the collection of personal data from smartphones (17–19). Finally, many apps find creative side channels to routinely extract data from people's phones (20, 21)—regardless of whether permission has been provided.

Here we evaluate whether individuals' Big Five personality trait levels can be predicted on the basis of six different classes of

Significance

Smartphones are sensor-rich computers that can easily be used to collect extensive records of behaviors, posing serious threats to individuals' privacy. This study examines the extent to which individuals' personality dimensions (assessed at broad domain and narrow facet levels) can be predicted from six classes of behavior: 1) communication and social behavior, 2) music consumption, 3) app usage, 4) mobility, 5) overall phone activity, and 6) day- and night-time activity, in a large sample. The cross-validated results show which Big Five personality dimensions are predictable and which specific patterns of behavior are indicative of which dimensions, revealing communication and social behavior as most predictive overall. Our results highlight the benefits and dangers posed by the widespread collection of smartphone data.

Author contributions: C.S. designed research; C.S., R.S., S.T.V., T.S., M.O., and T.U. performed research; C.S., Q.A., D.B., S.T.V., T.S., T.U., H.H., B.B., and M.B. contributed new reagents/analytic tools; C.S., Q.A., M.O., and T.U. analyzed data; C.S., Q.A., S.D.G., and G.M.H. wrote the paper; C.S., R.S., S.D.G., G.M.H., D.B., S.T.V., T.S., H.H., B.B., and M.B. improved manuscript; and H.H., B.B., and M.B. provided resources.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Data deposition: Data and code for this paper have been deposited in the Open Science Framework: https://osf.io/kqjhr/.

[1]To whom correspondence may be addressed. Email: stachl@stanford.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1920484117/-/DCSupplemental.

behavioral information collected via smartphones. Moreover, we examine which behaviors reveal most about each personality trait and how predictive each behavioral class is on average. Using sensor and log data from volunteers' smartphones, we extracted thousands of variables, categorized into six classes of daily behavior derived from previous research: 1) app usage (e.g., mean duration of gaming app usage), 2) music consumption (e.g., mean valence of played songs), 3) communication and social behavior (e.g., number of outgoing calls per day), 4) mobility behaviors (e.g., mean radius of gyration), 5) overall phone activity (e.g., number of unlock events per day), and 6) a higher-level behavioral class that captured the extent of daytime versus nighttime activity (e.g., outgoing calls at night). Together these six classes of behavior provided a broad sampling of the data that can easily be derived from smartphones and which may provide clues to individuals' personalities and allow for a robust investigation of our research question.

We assessed personality in terms of the Big Five dimensions, the most widely used and well-established system in psychological science for organizing personality traits (22–24). This taxonomy describes human personality in terms of five broad and relatively stable dimensions: openness, conscientiousness, extraversion, agreeableness, and emotional stability (22, 23), with each dimension subsuming a larger number of more specific facets. The Big Five have been found to have a strong genetic basis and to replicate across cultures and contexts (25–27).

Past studies have highlighted the promise of using smartphones to associate behavioral data with personality traits and other private attributes (28–39). A subset of these studies has used machine learning in analyses with the goal of predicting personality traits from behavioral measures (28–30, 38, 39). However, this subset of studies was subject to a number of key limitations, including the following: 1) focusing on just a single class of behavior or a small number of similar behaviors (e.g., communication behavior, refs. 31 and 39); 2) using small samples (28–30, 39); 3) being confined to the broad personality trait domains, not their more specific facets (28–30, 38, 39); 4) using classification instead of regression for the prediction of continuous personality scores (28, 29, 31); 5) likely overestimating model performance (28–30) (see ref. 31, for a discussion of the problem); 6) not providing enough information to reproduce findings (e.g., open data and materials, refs. 28–30, 38, and 39); and 7) not determining the relative effects of variables in the prediction models (28–30, 38).

To address these issues, we use smartphone sensing to gather behaviors from a wide variety of behavioral classes from a large sample, measure personality at both the domain and facet levels, train linear and nonlinear regression models (elastic net, random forest), properly evaluate our models out of sample using a (nested) cross-validated approach, and explore which behaviors are most predictive of personality overall and with respect to the individual personality domains and facets using interpretable machine learning and corrected significance tests. As a benchmark for the performance of our models, we compare the predictive performance with that of previous research using digital footprints from social media platforms (e.g., ref. 1).

Results

Personality Trait Prediction with Behavioral Patterns. Descriptive statistics can be found in *SI Appendix*, Tables S1 and S2, and in extensive detail on the project's website, accessible via the project repository (40). The results show that we successfully predicted levels of Big Five personality traits from behavioral patterns, derived from smartphone data, for more than half of the domains and facets (57% of all personality dimensions). In multiple instances both model types performed well above the baseline model (i.e., a model that constantly predicts the mean in the respective training set). Furthermore, our results suggest differences in how well the trait dimensions were predicted, as can be seen in Fig. 1 and in *SI Appendix*, Table S4 (e.g., sociableness most accurately and agreeableness not at all). The results also show that the nonlinear random forest models on average outperformed the linear elastic net models in both prediction performance and the number of successfully predicted criteria, hinting at the presence of nonlinear correlational structures in the data. Table 1 shows the top five most-important predictor variables per criterion. In Fig. 2 we provide a comprehensive visualization of all model results and effects of the behavioral classes. Fig. 2, *Top* shows the median prediction performance in R^2, and Fig. 2, *Upper Middle* shows the contribution and significance of a behavioral class by itself for the respective model (unique class importance). Fig. 2, *Lower Middle* shows the contribution of a behavioral class in the context of all other classes (combined class importance). Red circles indicate significant effects. In Fig. 2, *Bottom*, color-coded behavioral patterns ranked by variable importance are displayed across all models.

Here we report median prediction performances for all personality trait models, aggregated across the outer cross-validation folds. We report all metrics for both model types in *SI Appendix*, Table S4. In *SI Appendix*, Fig. S1 we also show exploratory predictor effects in accumulated local effect plots (ALEs). Additionally, we provide P values for the behavioral class effects, in *SI Appendix*, Table S5. For clarity and due to the model's superiority in prediction, we report performance metrics only for the random forest models in the text. However, results for both types of models, including plots, variable importance measures, and all exploratory single-predictor effects, are available on the project's website, accessible via the project's repository (40). In addition to results from predictive modeling, we also summarize findings from the interpretable machine-learning analyses. Below we describe which classes of behavior were significantly predictive for the respective personality dimension and provide some illustrative examples of single-variable effects, which should not be generalized beyond our sample. Finally, by refitting models on all combinations of the behavioral classes, we evaluate the average effect of each class for the prediction of personality trait dimensions. Data and code to reproduce all analyses are available in the project's repository (40).

Except for openness to imagination ($r_{\text{md}} = 0.19$, $r_{\text{sd}} = 0.13$), openness ($r_{\text{md}} = 0.29$, $r_{\text{sd}} = 0.11$) and its facets were successfully predicted in our dataset. With regard to facets, openness to aesthetics showed the highest median prediction performance ($r_{\text{md}} = 0.29$, $r_{\text{sd}} = 0.12$) and openness to actions ($r_{\text{md}} = 0.23$, $r_{\text{sd}} = 0.11$) the lowest, with openness to feelings ($r_{\text{md}} = 0.24$, $r_{\text{sd}} = 0.09$) and openness to ideas falling in between ($r_{\text{md}} = 0.24$, $r_{\text{sd}} = 0.11$). The top predictors in Table 1 and behavioral patterns in Fig. 2 suggest that music consumption also played a role in the prediction models for openness (e.g., quieter music), but this could not be confirmed by the unique and combined class-based variable importance scores in Fig. 2. Those scores suggest that overall patterns in app-usage behavior (e.g., increased camera usage, more photos, less usage of sports news apps) and for openness to actions communication and social behavior (e.g., ringing events, calls at night) were most important for the prediction of openness and its facets.

Conscientiousness ($r_{\text{md}} = 0.31$, $r_{\text{sd}} = 0.13$) was also successfully predicted above baseline, as were its facets, except for competence ($r_{\text{md}} = 0.19$, $r_{\text{sd}} = 0.11$). In terms of prediction performance, the facet love of order ranked first ($r_{\text{md}} = 0.31$, $r_{\text{sd}} = 0.13$), followed by sense of duty ($r_{\text{md}} = 0.29$, $r_{\text{sd}} = 0.10$), ambition ($r_{\text{md}} = 0.26$, $r_{\text{sd}} = 0.12$), discipline ($r_{\text{md}} = 0.22$, $r_{\text{sd}} = 0.12$),

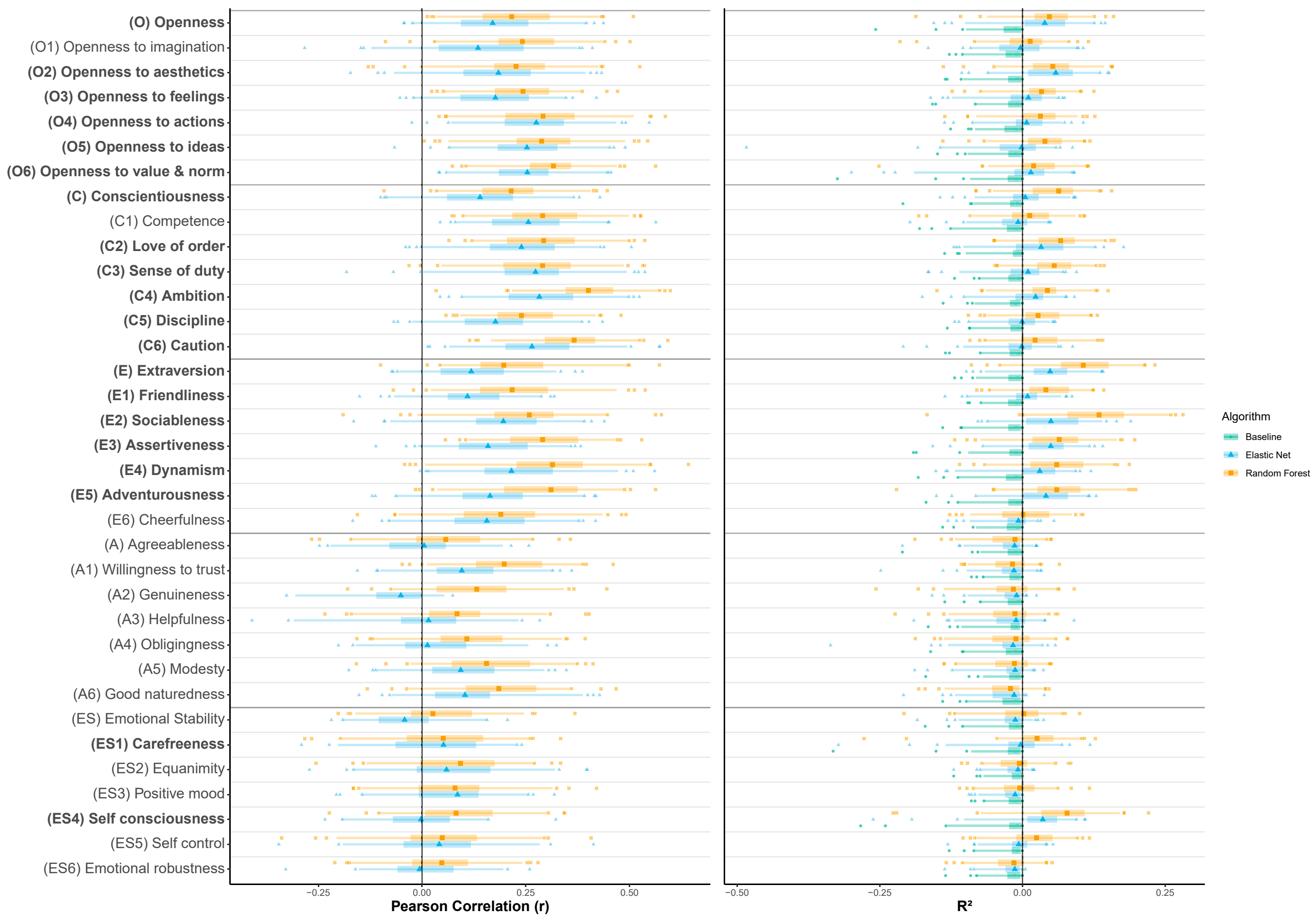

Fig. 1. Box and whisker plot of prediction performance measures from repeated cross-validation for each personality domain and facet. The middle symbol represents the median, boxes include values between the 25 and 75% quantiles, and whiskers extend to the 2.5 and 97.5% quantiles. Outliers are depicted by single points. Names of significant models are in boldface type. Figure is available at https://osf.io/kqjhr/, under a CC-BY4.0 license.

and caution (r_{md} = 0.20, r_{sd} = 0.12). Inspection of behavioral patterns and class importance indicators in Fig. 2 suggests that in the context of all other variables, predominantly variables related to overall phone activity (e.g., earlier first and last phone use per day), day and nighttime activity (e.g., less variable nightly duration of inactivity), and most unique app usage (e.g., increased usage of weather apps, timers, and checkup-monitoring apps) were especially important for the prediction of higher scores in the models of conscientiousness and its facets. Additionally, for the facets love of order and sense of duty, a very specific behavior was found to be important—the mean charge of the phone when it was disconnected from a charging cable. ALEs in *SI Appendix*, Fig. S1 suggest that in the context of all predictors higher average scores in love of order were predicted for charges above 60%.

With the exception of the cheerfulness facet (r_{md} = 0.16, r_{sd} = 0.12), the personality trait of extraversion (r_{md} = 0.37, r_{sd} = 0.09) and its facets were successfully predicted above baseline. Most notably, the facet of sociableness was predicted with the highest performance of all criteria (r_{md} = 0.40, r_{sd} = 0.10). Besides sociableness, the facets friendliness (r_{md} = 0.24, r_{sd} = 0.09), assertiveness (r_{md} = 0.29, r_{sd} = 0.11), dynamism (r_{md} = 0.29, r_{sd} = 0.10), and adventurousness (r_{md} = 0.29, r_{sd} = 0.11) were predicted above baseline. Behavioral patterns and class importance (unique and combined) in Fig. 2 suggest that variables related to communication and social behavior (e.g., higher mean number of outgoing calls per day, higher irregularity of all calls, higher mean number of WhatsApp uses per day) were important in the prediction of higher scores in the models of extraversion and its facets.

In the present analyses, the personality dimension of agreeableness could not be successfully predicted from the data, either on domain or on facet levels (r_{md} = 0.05, r_{sd} = 0.11).

For the personality dimension of emotional stability, only the facets of carefreeness (r_{md} = 0.22, r_{sd} = 0.10) and self-consciousness (r_{md} = 0.32, r_{sd} = 0.09) were predicted significantly. Behavioral patterns in Fig. 2 are rather distinct for the individual facets of emotional stability. Whereas communication and social behavior were significantly predictive for the facet self-consciousness (e.g., higher number of calls), the model of carefreeness did not show any significant effects at the class level.

In summary, all behavioral classes had some impact on the prediction of personality trait scores (as seen in Fig. 2). However, behaviors related to communication and social behavior and app usage showed as most significant in the models. This pattern can be discerned in Fig. 2. To estimate the average effect of each behavioral class on the prediction of personality trait dimensions overall (successfully and unsuccessfully predicted in the main analyses), we used a linear mixed model (details of the analysis are described in *Materials and Methods*). Results of the model show that communication and social behavior had the biggest impact on model performance on domain (β = 0.027, $\text{CI}_{95\%} = [.026, .028]$) and facet levels ($\beta$ = 0.019,

Table 1. Top five predictors per prediction model

Personality dimension	Top five predictors
O, openness	Daily mean length text messages | robust mean dur sports news apps | daily robust variation dur phone ringing | daily robust mean no. photos | robust mean dur sports news apps night
O2, openness to aesthetics	Robust mean dur sports news apps | daily mean no. photos | daily mean no. unique sports news apps | robust mean dur nightly sports news app | daily mean no. sports news apps
O3, openness to feelings	Excess music acousticness | daily mean no. unique sports news apps per week | robust variation dur shared transportation apps | daily robust variation in dur phone ringing | daily mean no. unique sports news apps
O4, openness to actions	Mean no. of phone ringing night | daily mean no. of ringing events | daily mean no. Google Maps | mean no. calls night | irregularity of phone ringing
O5, openness to ideas	Loudness fourth most listened song | robust mean dur sports news apps | daily SD no. of photos | robust mean dur *Süddeutsche Zeitung* (newspaper) | robust mean dur Samsung Notes
O6, openness to value and norm	Daily mean no. unique sports news week | daily mean no. Facebook | daily mean no. sports news | daily mean no. unique sports news weekend | daily mean no. Kicker (soccer news)
C, conscientiousness	Robust mean dur weather app night | daily SD sum interevent time | robust mean time last event | robust variation dur checkup monitoring apps | robust variation first event weekdays
C2, love of order	Daily SD sum interevent time | robust mean dur news-magazine apps | daily mean no. unique email apps | mean mean charge disconnection | robust variation dur TV-filmguide apps
C3, sense of duty	SD dur nightly downtime | robust mean time first event weekdays | robust variation time last event weekdays | robust mean dur Stadtwerke München Fahrinfo München (public transportation)
C4, ambition	Robust mean time first event | robust variation first event weekdays | robust mean time last event | robust variation time first event weekends | daily mean no. Google Playstore
C5, discipline	Robust variation time first event weekdays | robust mean time first event weekdays | robust mean dur weather apps night | robust variation time first event weekends | daily SD sum interevent time
C6, caution	Robust variation time last event weekdays | SD dur nightly downtime Sunday til Thursday | similarity contacts phone and messaging | robust variation time last event | mean music valence weekends
E, extraversion	Nightly mean no. phone ringing | nightly mean no. calls | daily mean no. outgoing calls | daily mean no. phone ringing | nightly mean no. outgoing calls
E1, friendliness	Daily mean no. phone ringing | irregularity of phone ringing weekend | daily SD no. incoming calls | daily robust variation sum dur phone ringing | daily SD sum dur incoming calls
E2, sociableness	Mean no. calls night | daily mean no. outgoing calls | mean no. phone ringing night | mean no. outgoing calls night | irregularity of phone ringing weekend
E3, assertiveness	Daily mean no. outgoing calls | daily mean no. contacts per week | daily mean no. contacts outgoing calls | daily mean no. contacts calls | mean no. calls night
E4, dynamism	Daily mean no. outgoing calls | mean no. phone ringing night | daily mean no. contacts outgoing calls | mean no. calls night | daily mean no. phone ringing
E5, adventurousness	Mean no. phone ringing night | mean no. calls night | irregularity of phone ringing | mean no. outgoing calls night | irregularity of calls
ES1, carefreeness	Daily mean no. Android-Email (app) | daily mean no. screen unlocks | robust variation dur system apps | robust variation dur strategy games | daily mean no. phone ringing
ES4, self-consciousness	Nightly mean no. calls | daily mean no. phone ringing | daily mean no. contacts calls | daily mean no. outgoing calls | daily mean no. contacts incoming calls

The top five most predictive features are shown for each successfully predicted personality dimension in the random forest models. The ranking is based on permutation feature importance and goes from left (high) to right (low). dur = duration.

$\mathrm{CI}_{95\%} = [.019, .020]$). App usage was second ($\beta_{\text{domains}} = 0.014$, $\mathrm{CI}_{95\%} = [.013, .015]$, $\beta_{\text{facets}} = 0.014$, $\mathrm{CI}_{95\%} = [.014, .015]$) followed by day and nighttime activity ($\beta_{\text{domains}} = 0.013$, $\mathrm{CI}_{95\%} = [.012, .014]$, $\beta_{\text{facets}} = 0.011$, $\mathrm{CI}_{95\%} = [.011, .012]$), overall phone activity ($\beta_{\text{domains}} = 0.006$, $\mathrm{CI}_{95\%} = [.005, .007]$, $\beta_{\text{facets}} = 0.004$, $\mathrm{CI}_{95\%} = [.004, .005]$), and music ($\beta_{\text{domains}} = 0.001$, $\mathrm{CI}_{95\%} = [.000, .002]$, $\beta_{\text{facets}} = 0.001$, $\mathrm{CI}_{95\%} = [.001, .002]$). The behavioral class of mobility was least important for the prediction of Big Five personality trait dimensions ($\beta_{\text{domains}} = -0.001$, $\mathrm{CI}_{95\%} = [-.002, -.001]$, $\beta_{\text{facets}} = -0.001$, $\mathrm{CI}_{95\%} = [-.001, .000]$). In *SI Appendix*, Fig. S2, we provide additional, exploratory results of a resampled greedy forward search analysis, indicating which combinations of behavioral classes were most predictive overall, in our dataset.

Discussion

The results presented here demonstrate that information about individuals' everyday behaviors detected from smartphone sensors and logs can be used to infer their Big Five personality trait dimensions. Specific classes of behavior (app usage, music consumption, communication and social behavior, mobility behavior, overall phone activity, daytime vs. nighttime activity) were distinctively informative about the different Big Five trait dimensions. Our models were able to predict personality on the broad domain level and the narrow facet level for openness, conscientiousness, and extraversion. For emotional stability, only single facets could be predicted above baseline. Finally, scores for agreeableness could not be predicted at all. The behavioral class of communication and social behavior was most important for the prediction of personality trait dimensions on average, but app usage and day and nighttime activity were also important[*]. We found performance levels across all significant models ($r_{range} = [0.20, 0.40]$) to be on average similar to those identified in a metaanalysis of previous studies predicting personality from digital footprints, which reported a mean effect size of $r = 0.34$ (1). As benchmarks for gauging these effect sizes,

[*] As can be seen in Fig. 2, in roughly half of the models the behavioral class communication and social was most important and, for the other half, app usage was most important.

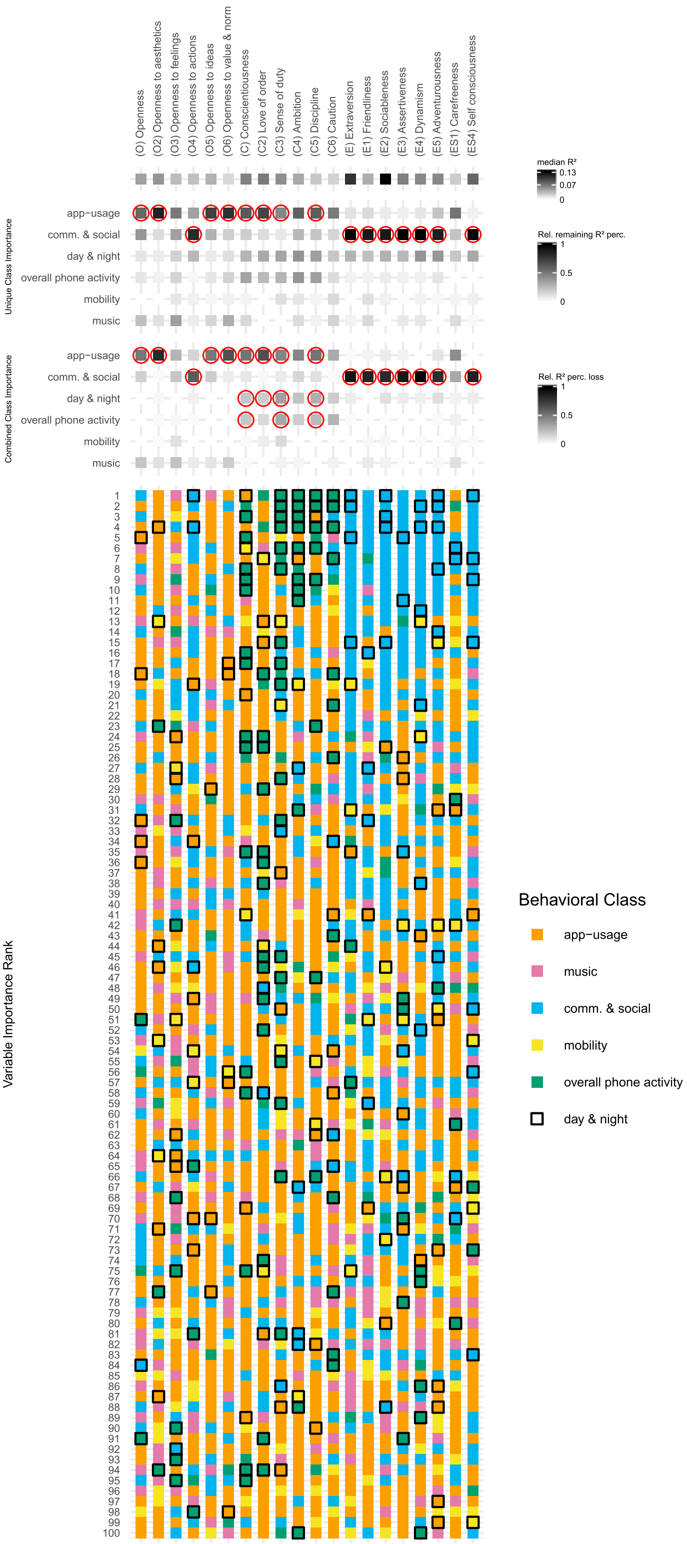

Fig. 2. (*Top*) Median prediction performance in R^2. (*Upper Middle*) Relative remaining performance when keeping variables of the respective class intact and permuting variables of other classes (unique class importance). (*Lower Middle*) Relative drop in performance when permuting variables of respective groups (combined class importance). Red circles indicate significant effects tested with the PIMP algorithm (41). (*Bottom*) Behavioral patterns of ranked permutation-based variable importance, color coded by class of behavior. Black frames indicate additional day–night dependency. Figure is available at https://osf.io/kqjhr/, under a CC-BY4.0 license.

consider that the highest median effect size ($r_{md} = 0.40$) is comparable to the tendency of people in a bad mood to be more aggressive than people in a good mood, and the smallest significant effect size is equal to the average reported effect size in personality psychology (42). These performance levels highlight the practical relevance of our results beyond significance.

The results here point to the breadth of behavior that can easily be obtained from the sensors and logs of smartphones and, more importantly, the breadth and specificity of personality predictions that can be made from the behavioral data so obtained. However, it is important to note that these findings are, if anything, a conservative estimate of what can be learned about people's personalities using information obtainable from their smartphones. Greater prediction accuracies would almost certainly be obtained when using more sensors (e.g., accelerometers, microphones, cameras; ref. 11); more log data (e.g., over longer time periods); content-level data (e.g., the content of texts, calls, emails, photos, videos, or all visible information on the screen; ref. 43); bigger, more diverse, and more representative samples (e.g., iPhone operating system [iOS] and Android users, nonwestern, educated, industrialized, rich, and democratic [WEIRD] samples; ref. 44); and by combining these data with other information about the user, derived from other sources (e.g., purchase histories, digital footprints from social media). Furthermore, models in this paper are still limited by the sparsity in the data (e.g., app usage), because some apps were used by only very few participants. Larger samples (e.g., as used in studies on personality social media use; ref. 6) could also allow for more accurate predictions.

As such, the present work serves as a harbinger of both the benefits and the dangers presented by the widespread use of behavioral data obtained from smartphones. On the positive side, obtaining behavior-based estimates of personality stands to open additional avenues of research on the causes and consequences of personality traits, as well as permitting consequential decisions (e.g., in personnel selection) to draw on behavioral data rather than estimates derived from self-report questionnaires, which are subject to a range of biases (e.g., responses biases, social desirability, different reference standards, memory limitations; refs. 45 and 46).

At the same time, we should not underestimate the potential negative consequences of the routine collection, modeling, and uncontrolled trade of personal smartphone data (20, 21, 47). For example, organizations and companies can obtain information about individuals' private traits (e.g., the Big Five personality traits), without the personality information ever being deliberately provided or explicitly requested (48). Mounting evidence suggests that these data can and are being used for psychological targeting to influence people's actions, including purchasing decisions (5, 47) and potentially voting behaviors, which are related to personality traits (49, 50).

Many commercial actors already collect a subset of the behavioral data that we have used in this work using publicly available applications (20). In academic settings, such data collection requires institutional review board (IRB) approval of the research study. However, current data protection laws in many nations do not adequately regulate data collection practices in the private sector. For example, in online real-time bidding on advertisements multiple actors exchange cross-device data to win bids to cater personalized ads to single users; this process is complex, happens within milliseconds, and is poorly understood outside of the industry (47). In such cases, once the data are collected from people's smartphones, the data's distribution seems to largely escape legislative oversight and legal enforcement (21, 47). This is the case even though legal frameworks against the routine collection of these data exist (e.g., the General Data Protection Regulation [GDPR] in the European Union; ref. 51) and reflects the growing asymmetry between one-click privacy permissions and the untraceable ways behavioral data from peoples' phones can wander.

Hence, a more differentiated choice with regard to the types of data and their intended usage should be given to users. For example, users should be made aware that behavioral data from phones are required for the completion of a specific task (e.g., finding a café); could be reused or sold to third parties, combined with other data; or used to create user models to make indirect predictions (e.g., personality, financial, credit scoring). In other words, it must be more obvious to consumers whether they are consenting to the measurement of their app use or to the automatic prediction of their private traits (e.g., personality).

Under most legislation, all of these actions are currently possible after initially providing the permission to access data on phones. One idea is for user data to have an automatic expiration date, after which data attributable to a unique identity must be deleted. Finally, the manifold techniques that online marketing companies use to link datasets of individuals to facilitate personalized ads (i.e., unique identifiers; ref. 47) could also be used to opt out of all advertisements and data-processing activities. Some variations of these suggestions are already implemented in the European Union's GDPR (51). We hope our findings stimulate further debate on the sensitivity of behavioral data from smartphones and how privacy rights can be protected at the individual (15) and aggregate levels (52).

A large portion of current economic and scientific progress depends on the availability of data about individuals' behaviors. The smartphone represents an ideal instrument to gather such information. Therefore, our results should not be taken as a blanket argument against the collection and use of behavioral data from phones. Instead, the present work points to the need for increased research at the intersection of machine learning, human computer interaction, and psychology that should inform policy makers. We believe that to understand complex social systems, while at the same time protecting the privacy of smartphone users, more sophisticated technical and methodological approaches combined with more dynamic and more transparent approaches to informed consent will be necessary (e.g., distributed privacy, federated learning, privacy nudges; refs. 53–56). These approaches could help balance the tradeoff between the collection of behavioral smartphone data and the protection of individual privacy rights, resulting in higher standards for consumers and industry alike.

Materials and Methods

Participants and Dataset. The dataset was collected in three separate studies as part of the PhoneStudy mobile sensing research project at the Ludwig-Maximilians-Universität München (LMU) (57). Parts of the data have been used in other publications (32, 33, 58, 59), but the joint dataset of common parameters has not been analyzed before. A total of 743 volunteers were recruited via forums, social media, blackboards, flyers, and direct recruitment, between September 2014 and January 2018 (33, 58, 59). All subjects participated willingly and provided informed consent prior to their participation in the study. Volunteers could withdraw from participation and demand the deletion of their data as long as their reidentification was possible. Dependent on the respective study (33, 58, 59), we provided different rewards for participation. Procedures for all studies were approved by the IRB of the Psychology Department at Ludwig-Maximilians-Universität München and have been conducted according to European Union laws. In *SI Appendix*, Table S3 we provide an overview of the datasets. We excluded data from volunteers with less than 15 d of logging data (29), no app usage (39), and missing questionnaire data (52). The final sample (n = 624) was skewed in favor of more educated (91% completed A levels, 20% had a university degree), younger participants (M = 23.56, SD = 6.63) and was not equally balanced with regard to gender (377 women, 243 men, and 4 with undisclosed gender).

Procedures. Study procedures were somewhat different across the three studies (33, 58, 59). However, in all three studies, Big Five personality trait levels were measured with the German version of the Big Five Structure Inventory (BFSI) (60) and naturalistic smartphone usage in the field was automatically recorded over a period of 30 d. The data were regularly transferred to our encrypted server using Secure Sockets Layer (SSL) encryption, when phones were connected to WiFi. In study 2, volunteers had to answer experience sampling questionnaires during the data collection period on their smartphones (59). Volunteers in studies 2 and 3 completed the demographic and BFSI personality questionnaires via smartphone at a convenient time (58). In cases where volunteers turned off location services, they were reminded to reactivate them. At the end of mobile data collection, volunteers were instructed to contact the research staff to receive compensation (studies 1 to 3) and to schedule a final laboratory session (study 2). More details about the procedures of the individual studies are available in the respective research articles (33, 58, 59).

Self-Reported Personality Measures and Demographics. Big Five personality dimensions were assessed with the German version of the BFSI (60). The test consists of 300 items and measures the Big Five personality dimensions (openness to experience, conscientiousness, extraversion, agreeableness, and emotional stability) on five domains and 30 facets. Participants indicated their agreement with items using a four-point Likert scale ranging from untypical for me to typical for me.

Additionally, we collected age, gender, highest completed education, and a number of other questionnaires that were used in other research projects. More information can be found in the respective online repositories and articles (33, 58, 59). Questionnaires were administered either via desktop computer (studies 1 and 2) or via smartphone (studies 2 and 3). We used the laboratory version scores from study 2 in this study. Descriptive statistics including confidence intervals of internal consistencies (α) are provided in *SI Appendix*, Table S1.

Behavioral Data from Smartphone Sensing. We used the PhoneStudy smartphone research app for Android to collect behavioral data from the volunteers' privately owned smartphones. This app has been continuously developed at the Ludwig-Maximilians-Universität München since September 2013.

Initially, activities were recorded in the form of time-stamped logs of events. Those events included calls, contact entries, texting, global positioning system (GPS) locations, app starts/installations, screen de/activations, flight mode de/activations, Bluetooth connections, booting events, played music, battery charging status, photo and video events, and connections to wireless networks (WiFi). Additionally, the character length of text messages and technical device characteristics were collected. Irreversibly hash-encoded versions of contacts and phone numbers were collected to enable us to measure the number of distinct contacts while preventing the possibility of reidentification. Information such as names, phone numbers, and contents of messages, calls, etc., was not recorded at any time.

Data Analysis. The final dataset consisted of 1,821 behavioral predictors and 35 personality criteria (five domains and 30 facets). Gender, age, and education were used solely for descriptive statistics and were not included as predictors in the models.

Variable extraction. In a first step, we extracted 15,692 variables from the raw dataset. The extracted variables roughly correspond to the aforementioned behavioral classes of app usage, music consumption, communication and social behavior, mobility, overall phone activity, and day- and nighttime dependency. Variables with regard to day and night dependency were not computed for music consumption behaviors. Besides common estimators (e.g., arithmetic mean, SD sum, etc.), we computed more complex variables containing information about the irregularity, the entropy, the similarity, and the temporal correlation of behaviors. These variables provided information about specific data types (e.g., mobility data) and were used for the quantification of behavioral structures within person and across time while avoiding more complex time-series models. The large amounts of data meant it was unfeasible to check for outliers manually, so we used robust estimators (e.g., Huber M Estimator; ref. 61) for most variables (except for call and messaging variables that were checked manually). Details about the calculation of variables and the full set of extracted variables and a detailed overview of all sensed data are provided in the project repository (40).

Machine learning. We fitted machine-learning models with an inner cross-validation loop (5-fold cross-validation [CV]) for preprocessing and hyperparameter tuning and an outer cross-validation loop (10 × 10-fold CV) for unbiased model evaluation. We compared the predictive performance of elastic net regularized linear regression models (62) with those of nonlinear tree-based random forest models (63) and a baseline model. The baseline model predicted the mean of the respective training set for all cases in a test set. We chose these standard models due to their ability to cope with $P >> N$ problems (i.e., few cases, many predictors). Furthermore, the usage of random forest models allowed us to include nonlinear predictor effects and high-dimensional interactions in the models.

We evaluated the predictive performance of the models based on the Pearson correlation (r) and the coefficient of determination (R^2). Specifically, we compared the predicted values from our models with the latent person-parameter trait estimates from the self-reported values of the personality trait measures. Because the personality scores in our analyses already represent latent trait scores, correlation measures were not adjusted for the reliability of the personality trait scales (all attenuated). Thus, the absolute size of the correlations is limited by the reliability of the personality trait measures. Disattenuated correlation coefficients are provided in *SI Appendix*, Table S5. We computed performance measures within each fold of the cross-validation procedure and averaged across all outer resampling folds within a single prediction model (e.g., for extraversion). To determine whether a model was predictive at all, we carried out t tests by comparing the R^2 measures of the random forest model with those of the baseline model. The t tests were based on 10-times repeated 10-fold cross-validation and used a variance correction to specifically address the dependence structure of cross-validation experiments (64). All comparisons were adjusted for multiple comparisons (n = 35) via Holm correction. Significant prediction models ($\alpha = 0.05$) are marked in boldface type in Fig. 1.

In addition to measures of predictive performance, we used interpretable machine-learning techniques with significant models to gain insights into our models' inner workings. Specifically, we used permutation strategies to determine the unique contribution of the respective behavioral class and the importance of a class within the context of all other classes. These effects were also tested for significance (41) and adjusted for multiple comparisons.

To determine which of the behavioral classes was the most important overall for the prediction of Big Five personality traits, we performed an additional resampling analysis: 1) We created predictor sets with all possible combinations of subsets of the six behavioral classes (2^6 = 64); 2) we created 100 resampling folds of the complete dataset (10-times repeated 10-fold cross-validation; train and test data splits remained the same across all combinations); 3) for each of these combinations in all folds ($64 \times 100 = 6{,}400$), we fitted (on training data) and evaluated (on test data) models to predict each personality criterion (30 facets or 5 domains = 30 or 5 R^2 coefficients); 4) we averaged R^2 across all personality criteria, within each fold of a combination (100 mean R^2 values); and 5) we used two maximum-likelihood linear mixed models (domains vs. facets) with the mean R^2 as the outcome variable, the resampling iteration as the random factor (fold 1 to 100), and the behavioral classes (dummy encoded) as fixed factors. This procedure allowed us to determine the effects of each behavioral class on the average prediction performance across all personality trait dimensions. P values in the linear mixed models were adjusted for multiple testing with the Holm method. All procedures were performed on domain and facet levels, separately. Further details about preprocessing, the modeling procedures, and the performance metrics are available in *SI Appendix* and in the project's repository (40).

Software. Due to the high computational load of the machine-learning analyses, we parallelized the computations on the Linux Cluster of the LRZ-Supercomputing Center, in Garching, near Munich, Germany. For computations on the cluster, R-version 3.5.0 was used (65). We used R 3.5.2 for all other analyses (65). We used the fxtract package (66) for variable extraction from the raw data. Furthermore, we used the mlrCPO (67) and caret (68) packages for preprocessing. For machine learning we used the mlr (69), glmnet (70), iml (71), and ranger (72) packages.

Open data and materials and additional resources. We provide the dataset and the code for variable extraction, preprocessing, and modeling in the project's repository (40). Raw data files cannot be provided (due to unsolved privacy implications); full reproducibility is possible for the analyses but not for preprocessing and variable extraction. In the repository, we link to the interactive project website where readers can find an exhaustive data dictionary, additional methodological descriptions, references, and results for all models in much greater detail. This paper is based on a preprint (73).

ACKNOWLEDGMENTS. We thank L. Sust, F. Bemmann, M. Schiele, A. Dietl, V. Gebhardt, S. Huber, M. Fröhlich, M. Metz, F. Lehmann, P. A. Vu, D. Becker, J. Kaiser, P. Ehrich, A. De Luca, and M. Lamm for help with recruitment, programming, and app testing. We thank the Schuhfried GmbH for providing the BFSI. We thank F. Pargent, J. Rauthmann, F. Schönbrodt, and A. Bender for insightful comments on earlier versions of the manuscript. We thank N. Zschäck for her help with data visualization. Finally, we thank the reviewers for their constructive suggestions. This project was partially funded by the Bavarian State Ministry of Science and the Arts in the framework of the Center Digitisation.Bavaria, a Google research grant, the LMU-excellence initiative, the German Federal Ministry of Education and Research Grant 01IS18036A, and the National Science Foundation Award SES-1758835.

1. M. Settanni, D. Azucar, D. Marengo, Predicting individual characteristics from digital traces on social media: A meta-analysis. *Cyberpsychol. Behav. Soc. Netw.* **21**, 217–228 (2018).
2. D. J. Ozer, V. Benet-Martínez, Personality and the prediction of consequential outcomes. *Annu. Rev. Psychol.* **57**, 401–421 (2006).
3. B. W. Roberts, N. R. Kuncel, R. Shiner, A. Caspi, L. R. Goldberg, The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspect. Psychol. Sci.* **2**, 313–45 (2007).
4. C. J. Soto, How replicable are links between personality traits and consequential life outcomes? The life outcomes of personality replication project. *Psychol. Sci.* **30**, 711–727 (2019).
5. S. C. Matz, M. Kosinski, G. Nave, D. J. Stillwell, Psychological targeting as an effective approach to digital mass persuasion. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 12714–12719 (2017).
6. W. Youyou, M. Kosinski, D. Stillwell, Computer-based personality judgments are more accurate than those made by humans. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 1036–1040 (2015).
7. International Telecommunication Union, Measuring the information society report 2018. *ITU Publ.* **1**, 2–18 (2018).
8. G. M. Harari *et al.*, Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspect. Psychol. Sci.* **11**, 838–854 (2016).
9. G. M. Harari, S. D. Gosling, R. Wang, A. T. Campbell, Capturing situational information with smartphones and mobile sensing methods. *Eur. J. Pers.* **29**, 509–511 (2015).
10. G. Miller, The smartphone psychology manifesto. *Perspect. Psychol. Sci.* **7**, 221–237 (2012).
11. S. Servia-Rodríguez *et al.*, "Mobile sensing at the service of mental well-being: A large-scale longitudinal study" in *26th International World Wide Web Conference, WWW 2017* (International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 2017), pp. 103–112.
12. K. K. Rachuri *et al.*, "EmotionSense: A mobile phones based adaptive platform for experimental social psychology research" in *UbiComp'10 - Proceedings of the 2010 ACM Conference on Ubiquitous Computing* (Association for Computing Machinery, New York, NY, 2010), pp. 281–290.
13. S. Saeb, E. G. Lattie, S. M. Schueller, K. P. Kording, D. C. Mohr, The relationship between mobile phone location sensor data and depressive symptom severity. *PeerJ* **4**, e2537 (2016).
14. S. Thomée, Mobile phone use and mental health. A review of the research that takes a psychological perspective on exposure. *Int. J. Environ. Res. Publ. Health* **15**, 2692 (2018).
15. G. M. Harari, A process-oriented approach to respecting privacy in the context of mobile phone tracking. *Curr. Opin. Psychol.* **31**, 141–147 (2019).
16. C. M. Gray, Y. Kou, B. Battles, J. Hoggatt, A. L. Toombs, "The dark (patterns) side of ux design" in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, CHI '18* (Association for Computing Machinery, New York, NY, 2018), pp. 1–14.
17. S. Barocas, H. Nissenbaum, "On notice: The trouble with notice and consent" in *Proceedings of the Engaging Data Forum: The First International Forum on the Application and Management of Personal Electronic Information* (2009). https://ssrn.com/abstract=2567409. Accessed 7 July 2020.
18. A. P. Felt *et al.*, "Android permissions: User attention, comprehension, and behavior" in *Proceedings of the Eighth Symposium on Usable Privacy and Security, SOUPS '12* (Association for Computing Machinery, New York, NY, 2012).
19. P. Wijesekera *et al.*, "The feasibility of dynamically granted permissions: Aligning mobile privacy with user preferences" in *Proceedings - IEEE Symposium on Security and Privacy* (Institute of Electrical and Electronics Engineers Inc., New York, NY, 2017), pp. 1077–1093.
20. J. Reardon *et al.*, "50 ways to leak your data: An exploration of apps' circumvention of the android permissions system" in *28th USENIX Security Symposium (USENIX Security 19)* (USENIX Association, Santa Clara, CA, 2019), pp. 603–620.
21. J. Valentino-DeVries, N. Singer, M. Keller, A. Krolik, Your apps know where you were last night, and they're not keeping it secret. *New York Times*, 10 December 2018.
22. L. R. Goldberg, An alternative "description of personality": The big-five factor structure. *J. Pers. Soc. Psychol.* **59**, 1216–1229 (1990).
23. R. R. McCrae, O. P. John, An introduction to the five-factor model and its applications. *J. Pers.* **60**, 175–215 (1992).
24. B. De Raad, The big five personality factors: The psycholexical approach to personality (Hogrefe & Huber Publishers, Seattle, WA, 2000).
25. J. C. Loehlin, R. R. McCrae, P. T. Costa, O. P. John, Heritabilities of common and measure-specific components of the big five personality factors. *J. Res. Pers.* **32**, 431–453 (1998).
26. P. Costa Jr, A. Terracciano, R. R. McCrae, Gender differences in personality traits across cultures: Robust and surprising findings. *J. Pers. Soc. Psychol.* **81**, 322–331 (2001).
27. C. M. Ching *et al.*, The manifestation of traits in everyday behavior and affect: A five-culture study. *J. Res. Pers.* **48**, 1–16 (2014).

28. G. Chittaranjan, J. Blom, D. Gatica-Perez, Mining large-scale smartphone data for personality studies. *Personal Ubiquitous Comput.* **17**, 433–450 (2013).
29. Y. A. De Montjoye, J. Quoidbach, F. Robic, A. Pentland, "Predicting personality using novel mobile phone-based metrics" in *Proceedings of the 6th International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction, SBP'13*, A. M. Greenberg, W. G. Kennedy, N. D. Bos, Eds. (Springer-Verlag, Berlin/Heidelberg, Germany, 2013), pp. 48–55.
30. W. Wang *et al.*, Sensing behavioral change over time: Using within-person variability features from mobile sensing to predict personality traits. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2**, 1–21 (2018).
31. B. Mønsted, A. Mollgaard, J. Mathiesen, Phone-based metric as a predictor for basic personality traits. *J. Res. Pers.* **74**, 16–22 (2018).
32. G. M. Harari *et al.*, Sensing sociability: Individual differences in young adults' conversation, calling, texting, and app use behaviors in daily life. *J. Pers. Soc. Psychol.*, 10.1037/pspp0000245 (2019).
33. C. Stachl *et al.*, Personality traits predict smartphone usage. *Eur. J. Pers.* **31**, 701–722 (2017).
34. C. Montag *et al.*, Smartphone usage in the 21st century: Who is active on WhatsApp? *BMC Res. Notes* **8**, 331 (2015).
35. C. Montag *et al.*, Correlating personality and actual phone usage: Evidence from psychoinformatics. *J. Indiv. Differ.* **35**, 158–165 (2014).
36. P. Ai, Y. Liu, X. Zhao, Big Five personality traits predict daily spatial behavior: Evidence from smartphone data. *Pers. Indiv. Differ.* **147**, 285–291 (2019).
37. L. Alessandretti, S. Lehmann, A. Baronchelli, Understanding the interplay between social and spatial behaviour. *EPJ Data Sci.* **7**, 36 (2018).
38. N. K. Kambham, K. G. Stanley, S. Bell, "Predicting personality traits using smartphone sensor data and app usage data" in *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference, IEMCON 2018* (IEEE, New York, NY, 2019), pp. 125–132.
39. N. Gao, W. Shao, F. D. Salim, Predicting personality traits from physical activity intensity. *Computer* **52**, 47–56 (2019).
40. C. Stachl *et al.* Repository: Predicting personality from patterns of behavior collected with smartphones. Open Science Framework. https://osf.io/kqjhr/. Deposited 18 June 2019.
41. A. Altmann, L. Toloşi, O. Sander, T. Lengauer, Permutation importance: A corrected feature importance measure. *Bioinformatics* **26**, 1340–1347 (2010).
42. D. C. Funder, D. J. Ozer, Evaluating effect size in psychological research: Sense and nonsense. *Adv. Methods Pract. Psychol. Sci.* **2**, 156–168 (2019).
43. B. Reeves, T. Robinson, N. Ram, Time for the human screenome project. *Nature* **577**, 314–317 (2020).
44. J. Henrich, S. J. Heine, A. Norenzayan, The weirdest people in the world? *Behav. Brain Sci.* **33**, 61–83 (2010).
45. P. M. Podsakoff, S. B. MacKenzie, N. P. Podsakoff, Sources of method bias in social science research and recommendations on how to control it. *Annu. Rev. Psychol.* **63**, 539–569 (2012).
46. Y. V. Vaerenbergh, T. D. Thomas, Response styles in survey research: A literature review of antecedents, consequences, and remedies. *Int. J. Publ. Opin. Res.* **25**, 195–217 (2013).
47. ICO, "Update report into adtech and real time bidding" (Tech. Rep., Information Commissioner's Office, UK, 2019). https://ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf. Accessed 7 July 2020.
48. M. Kosinski, D. Stillwell, T. Graepel, Private traits and attributes are predictable from digital records of human behavior. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5802–5805 (2013).
49. A. Roets, I. Cornelis, A. Van Hiel, Openness as a predictor of political orientation and conventional and unconventional political activism in western and eastern Europe. *J. Pers. Assess.* **96**, 53–63 (2014).
50. R. L. Bach *et al.*, Predicting voting behavior using digital trace data. *Soc. Sci. Comput. Rev.*, 10.1177/0894439319882896 (2019).
51. European Parliament, REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation, 2016).
52. Y. A. de Montjoye *et al.*, On the privacy-conscientious use of mobile phone data. *Sci. Data* **5**, 180286 (2018).
53. S. C. Matz, R. E. Appel, M. Kosinski, Privacy in the age of psychological targeting. *Curr. Opin. Psychol.* **31**, 116–121 (2020).
54. B. McMahan, E. Moore, D. Ramage, S. Hampson, B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data" in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research*, A Singh, J Zhu, Eds. (PMLR, Fort Lauderdale, FL, 2017), **vol. 54**, pp. 1273–1282.
55. J. Hong, The privacy landscape of pervasive computing. *IEEE Pervasive Comput.* **16**, 40–48 (2017).
56. H. Almuhimedi *et al.*, "Your location has been shared 5,398 times! A field study on mobile app privacy nudging" in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15* (Association for Computing Machinery, New York, NY, 2015), pp. 787–796.
57. C. Stachl *et al.*, Data from "The PhoneStudy project." Open Science Framework. https://osf.io/ut42y/. Accessed 7 July 2020.
58. R. Schoedel *et al.*, Digital footprints of sensation seeking. *Z. Psychol.* **226**, 232–245 (2018).
59. T. Schuwerk, L. J. Kaltefleiter, J. Q. Au, A. Hoesl, C. Stachl, Enter the wild: Autistic traits and their relationship to mentalizing and social interaction in everyday life. *J. Autism Dev. Disord.* **49**, 4193–4208 (2019).
60. M. Arendasy, *BFSI: Big-Five Struktur-Inventar (Test & Manual)* (Schuhfried GmbH, Mödling, Austria, 2009).
61. P. J. Huber, "Robust statistics" in *Wiley Series in Probability and Statistics* (John Wiley & Sons, Inc., 1981).
62. H. Zou, T. Hastie, Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B Stat. Methodol.* **67**, 301–320 (2005).
63. L. Breiman, Random forests. *Mach. Learn.* **45**, 5–32 (2001).
64. R. R. Bouckaert, E. Frank, "Evaluating the replicability of significance tests for comparing learning algorithms" in *Advances in Knowledge Discovery and Data Mining*, H Dai, R Srikant, C Zhang, Eds. (Springer Berlin Heidelberg, Berlin/Heidelberg, Germany, 2004), pp. 3–12.
65. R Core Team, R Development Core Team, R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, 2018).
66. Q. Au, C. Stachl, R. Schoedel, T. Ullmann, A. Hofheinz, *fxtract: Feature Extraction from Grouped Data, R Package Version 0.9.2* (The Comprehensive R Archive Network, 2019).
67. M. Binder, *mlrCPO: Composable Preprocessing Operators and Pipelines for Machine Learning, R Package Version 0.3.4* (The Comprehensive R Archive Network, 2018).
68. M. Kuhn *et al.*, *caret: Classification and Regression Training, R package version 6.0-79* (The Comprehensive R Archive Network, 2018).
69. B. Bischl *et al.*, mlr: Machine learning in R. *J. Mach. Learn. Res.* **17**, 1–5 (2016).
70. J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
71. C. Molnar, B. Bischl, G. Casalicchio, iml: An R package for interpretable machine learning. *JOSS* **3**, 786 (2018).
72. M. N. Wright, A. Ziegler, ranger: A fast implementation of random forests for high dimensional data in C++ and R. *J. Stat. Software* **77**, 1–17 (2017).
73. C. Stachl *et al.*, Behavioral patterns in smartphone usage predict big five personality traits. https://doi.org/10.31234/osf.io/ks4vd (12 June 2019).