European Journal of Personality, Eur. J. Pers. **34**: 632–648 (2020)
Published online 21 September 2020 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/per.2305

Using Big Data and Machine Learning in Personality Measurement: Opportunities and Challenges

LEO ALEXANDER III*, EVAN MULFINGER and FREDERICK L. OSWALD

Department of Psychological Sciences, Rice University, Houston, TX USA
leo.alexander@rice.edu

Abstract: This conceptual paper examines the promises and critical challenges posed by contemporary personality measurement using big data. More specifically, the paper provides (i) an introduction to the type of technologies that give rise to big data, (ii) an overview of how big data is used in personality research and how it might be used in the future, (iii) a framework for approaching big data in personality science, (iv) an exploration of ideas that connect psychometric reliability and validity, as well as principles of fairness and privacy, to measures of personality that use big data, (v) a discussion emphasizing the importance of collaboration with other disciplines for personality psychologists seeking to adopt big data methods, and finally, (vi) a list of practical considerations for researchers seeking to move forward with big data personality measurement and research. It is expected that this paper will provide insights, guidance, and inspiration that helps personality researchers navigate the challenges and opportunities posed by using big data methods in personality measurement. © 2020 European Association of Personality Psychology

Key words: big data; machine learning; personality measurement

Today's technologies, and the big data that they give rise to, have dramatically infused our everyday lives and continue to transform society in many domains, such as through social media, marketing, online education, and voting, and through continued developments in areas such as autonomous vehicles, personalized medicine, and automation in the workforce. This ongoing transformation not only reflects the inexorable impact of technology on society but also promises to benefit science by revealing unique undiscovered aspects of individuals' lives and personality. As researchers continue to investigate personality within the arena of big data and artificial intelligence (AI), knowledge of an interrelated and evolving system of benefits and challenges along ethical, legal, and scientific fronts is beginning to accrue. Furthermore, as we continue to think about and investigate personality in the domains of big data and machine learning, many ways of viewing big data personality measurement are likely to be useful, for example, (i) as a *cultural factor* that allows socio-technological contexts, relationships, and communications to develop in ways not possible just decades ago, creating new social behaviours and norms relevant to personality; (ii) as an *advisor and trainer* that reciprocates between behaviour and adaptive self-management interventions in a real-time system reflective of personality; (iii) as an *analyser and informant* that gathers, analyses, summarizes, and reports personality-relevant information to decision makers (e.g. teachers, supervisors, groups, and teams); and (iv) as a *scientific resource* that ideally provides personality-relevant data to support and improve future research and application efforts.

All these potential applications of big data for measuring and analysing personality and its outcomes pose new and interrelated technological, legal, and ethical questions and challenges for personality researchers and practitioners. Fortunately, the science of personality is ready to take on this future, given its long and rich history of developing, implementing, and evaluating measures of personality. In this context, the current paper explores the research opportunities and challenges associated with big data personality measurement, connecting future promise to acquired research wisdom by building on traditional psychometric concepts of reliability and validity, as well as principles of fairness and privacy.

Out of necessity, our overview will be general, given that personality research is just beginning to employ and investigate questions involving big data, AI, and machine learning, a fast-moving arena that opens many exciting doors for personality research. First, we will introduce technologies that have allowed for the collection of big data relevant to personality. Second, we will provide a brief overview of the small-but-growing body of existing big data personality research. Third, we will offer an organizing framework for thinking about big data in personality science. Fourth, we will re-examine traditional psychometric evaluations of reliability and validity, as well as fairness and privacy, within the modern context of big data algorithms as applied to personality-relevant big data. Fifth, we reflect on the critical importance of multidisciplinary collaboration for personality psychologists seeking to adopt big data methods. Finally, we will list key considerations that can inform future big data personality measurement and research.

*Correspondence to: Leo Alexander III, Department of Psychological Sciences, Rice University, 6100 Main St. MS25, Houston, TX 77005, USA. E-mail: leo.alexander@rice.edu

Handling editor: John Rauthmann
Received 11 November 2019
Revised 25 August 2020, Accepted 25 August 2020

DIGITAL TRACES OF BEHAVIOUR

Modern digital technologies have given rise to many potential sources of personality-relevant big data (see Woo, Tay, Jebb, Ford, & Kern, 2020). Smartphones and other personal electronic devices contain a variety of sensors (e.g. microphones, cameras, light sensors, accelerometers, and proximity sensors) and data logs (e.g. call logs, text messaging logs, web browser logs, and application use logs), both of which provide a rich source of longitudinal behavioural data (Harari et al., 2016). Bluetooth wireless data and GPS navigation features on cell phones can track where you are, when, and with whom—all of which may be revealing of personality (Mønsted, Mollgaard, & Mathiesen, 2018). Access to buildings, vehicles, and other secure locations increasingly rely on technologies that record one's specific whereabouts throughout the day (e.g. geographic location, building location via electronic access cards/codes, or video surveillance), which along with the associated time, and perhaps the people with which one was associated (and their data), might yield personality-relevant information. Online downloads and cloud-streaming services store data on whether you are listening to Mozart or Metallica, which can be translated into musical preferences that, in turn, inform personality (Rentfrow & Gosling, 2003). Digital financial transactions such as online purchases, debit card, credit card, and payment app transactions continue to increase in popularity (Foster, Schuh, & Zhang, 2013) and can be revealing of personality, as was demonstrated by Gladstone, Matz, and Lemaire (2019), who used machine learning techniques to predict personality from individuals' bank records. Although the accuracy of their predictions were small for the broad Big Five personality traits (*rs* = .12 to .19), they found larger correlations between their model's predictions and scores on survey-based measures of the more narrow and substantively relevant traits of materialism and self-control (*rs* = .33 and .26, respectively).

Technological advances not only offer new sources of personality-relevant data but also allow us to collect those data in new and more intensive ways. The trail for considering these data has been blazed by personality and methodological researchers who have focused strongly on investigating within-person variability that is configural (e.g. Molenaar & Campbell, 2009), context-driven (e.g. Fleeson, 2004), and time-driven (Hamaker, Nesselroade, & Molenaar, 2007). We are thus witnessing a dramatic resurgence of personality research focused on empirical tests of reliable patterns of within-person variation that is either related to or directly reflective of different personality traits (see the review by Beck & Jackson, 2019a), for example, examining relatively fast change, using daily diaries in clinical assessment (Zimmermann et al., 2019); momentary assessments of personality-performance relationships at work (Ilies & Judge, 2002; Judge, Simon, Hurst, & Kelley, 2014; Minbashian, Wood, & Beckmann, 2010); longer term life-event-driven changes in personality (Bleidorn, Hopwood, & Lucas, 2018); and developmental personality changes in making school-to-work transitions (Lüdtke, Roberts, Trautwein, & Nagy, 2011).

Wearable devices are particularly relevant to within-person variation in behaviour, as they are not only helpful to a user's understanding and self-management, but they also open up great possibilities for intervening in a more timely and tailored manner upon people at work, at school, or in daily life (Ihsan & Furnham, 2018; Mardonova & Choi, 2018). Personality may influence the type of intervention selected as well as how the intervention is responded to; and the interventions in turn may induce new habits, if not longer lasting personality change. For example, in the sports and medical domain, digital activity trackers (e.g. Fitbit and Apple Watch) are being marketed to fitness enthusiasts as well as anyone seeking to improve their health (Aroganam, Manivannan, & Harrison, 2019). The resultant fitness data may shed light, in aggregate groups and for individuals, on whether personality traits predict health decisions, processes, and outcomes such as (i) the choice to use these devices; (ii) the nature and difficulty of self-set health goals and; (iii) the process of how and when people revise, maintain, and attain those goals effectively.

Among the many advances in digital technologies, social media platforms stand out as particularly rich sources of big data for personality researchers. Given the large number of people who now regularly use social media networks (Lenhart et al., 2015; Perrin & Anderson, 2019), they often reflect not only extremely large samples and wide-ranging demographics (race/ethnicity, gender, age, geography, and culture) but also intensive longitudinal sampling, with digital trace data often accumulating multiple times per day. Perhaps it is no wonder, then, as we will discuss later, much of the existing big data personality research has relied upon data supplied by social media platforms.

THE CURRENT STATE OF PERSONALITY MEASUREMENT USING BIG DATA

Initial research efforts using trace data for personality measurement were conducted primarily in the field of computer science, where predictive models were applied to big data to predict scores on measures of the Five Factor model of personality (McCrae & John, 1992). This burgeoning body of research has identified many big data predictors of personality as extracted from a variety of digital sources, such as information from Facebook (Bachrach, Kosinski, Graepel, Kohli, & Stillwell, 2012; Golbeck, Robles, & Turner, 2011; Schwartz et al., 2013; Sumner, Byers, & Shearing, 2011; Wald, Khoshgoftaar, & Sumner, 2012; Youyou, Kosinski, & Stillwell, 2015; Youyou, Stillwell, Schwartz, & Kosinski, 2017), Twitter (e.g. Golbeck, Robles, Edmondson, & Turner, 2011; Quercia, Kosinski, Stillwell, & Crowcroft, 2011; Sumner, Byers, Boochever, & Park, 2012), Flickr (e.g. Yan et al., 2015), smartphone data (e.g. Chittaranjan, Blom, & Gatica-Perez, 2013), and personal weblogs (e.g. Iacobelli, Gill, Nowson, & Oberlander, 2011). Researchers continue to explore a variety of trace data from these and many other sources, including social media profile information, images, various user logs, messages, posts, and Facebook and Twitter 'likes'.

DOI: 10.1002/per

For example, the myPersonality project (Stillwell & Kosinski, 2020) developed an app where, until it closed in 2012, an astonishing six million Facebook users had taken at least one of a suite of online psychological tests available, including self-report and peer-report Big Five personality measures. Of those respondents, 40% agreed to share their anonymized data with the myPersonality project. Until the authors stopped sharing myPersonality data in 2018, these data were made available to other researchers, leading to dozens of peer-reviewed journal publications and greatly expanding the body of big data personality research (see mypersonality.org for a reflection on the project and a selected bibliography).

In addition to such large-scale studies, two recent meta-analyses have been conducted to estimate correlations between social media digital trace data and traditional Big Five questionnaires (Azucar, Marengo, & Settanni, 2018; Settanni, Azucar, & Marengo, 2018). Meta-analytic estimates suggest that correlations between digital trace data and measures of personality range from .29 to .40 across the Big Five personality traits, perhaps a reasonable level of convergence given the noisiness of trace data. Comparing these meta-analytic findings with a meta-analysis of the accuracy of human ratings of personality based on social media data (Tskhay & Rule, 2014), computer algorithms were at least more consistent than human raters (Youyou et al., 2015), if not more accurate given the absence of a gold standard for measures of personality.

Although using machine learning and big data to predict scores on the Big Five is useful, there is a much more extensive and continuing need to understand convergent and discriminant patterns involving machine learning models and personality-relevant data within a much larger nomological net of psychological constructs (Campbell & Fiske, 1959). Related to this need, additional research could also pursue three key questions about big data: (i) whether they yield unique personality-relevant variance; (ii) whether they provide incremental prediction of school, organizational, and life outcomes above traditional Big Five measures (Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007); and (iii) whether this incremental prediction can be attributed to personality and/or to other constructs.

As we have briefly described, research has already begun to make concerted efforts in predicting personality from big data; but whether *additional* personality insights actually result from big data seems to remain fertile investigative soil, for example, how have the clustering and/or predictive algorithms involving big data improved our understanding of personality? And how does an improved understanding of personality translate into people's functioning and well-being in their daily lives at work, school, and home? Before we partially address these questions in the current paper, let us first place some useful contours around the meaning of big data in personality research.

A FRAMEWORK FOR BIG DATA IN THE PERSONALITY CONTEXT

So many definitions of big data exist that they could themselves populate a big data set. Typically, these definitions do not focus on personality or any psychological characteristics but instead are broadly sensitive to unique characteristics of the data arising from the aforementioned digital technologies, quite often employing several Vs in doing so, for example, the historically well-known *volume*, *variety*, and *veracity* of big data proposed nearly 20 years ago (Laney, 2001); but also *visualization*, *variability*, *vulnerability*, *visibility*, *vagueness*—and still other Vs (Storey & Song, 2017). Such definitional terms are important for their useful abstractions, yet as any personality researcher dealing with big data would tell you, the devil of understanding big data lies in the important details associated with any particular application and setting. For example, technology interacts with humans, in addition to recording their behaviour (e.g. multimedia platforms use algorithms to suggest media as well as record a user's preferences in media), and thus, the influences of AI technology and human personality will be reciprocal. We say this knowing that technologies will always keep improving in ways that will better evoke, detect, and reflect personality (e.g. more effective AI, new data analytic techniques, faster processing speed, and greater storage capacity; Gandomi & Haider, 2015).

In the context of these changes, our perspective on big data relevant to personality research leans towards the practical: definitions and frameworks for big data are useful not only for their content but also as a tool for cultivating a growing community of psychological and interdisciplinary researchers who engage in a wide range of projects and applications involving personality-oriented big data. Keeping this perspective firmly in mind, we selected the framework proposed by De Mauro, Greco, and Grimaldi (2015) as a general and practical guiding framework for approaching big data (versus traditional data sets). These authors surveyed a wide range of definitions of big data and identified four common and distinct themes.

Information

Most frequently, definitions of big data include *information* or the size and structure of the data itself. This is related to the aforementioned 'Vs' and it is worth pointing out that in the history of psychological science, Cattell's *data box* (Cattell, 1946) parallels this information component of big data (Adjerid & Kelley, 2018). The data box is conceptual and reflects three dimensions of information—people × variables, people × occasions, and variables × occasions—where diverse types of rich within-person and between-person big data can populate this matrix (e.g. experience sampling methodologies can vary in their sampling rates in terms of people, variables, and occasions; Haqiqatkhah & Tuerlinckx, 2019). Also historically, Cronbach and colleagues' generalizability theory is an important-yet-underappreciated extension of the data box (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Cronbach, Rajaratnam, & Gleser, 1963) to estimate the variance along an infinite variety of dimensions (known as *facets* in generalizability theory), such as units, treatments, operations, and settings, called UTOS (Cronbach, 1982). The collection, analysis, and interpretation of personality-oriented big data can usefully benefit from this historical foundation of thinking

DOI: 10.1002/per

that comes with Cattell's data box and Cronbach et al.'s generalizability theory, because it lays out an important framework for research projects, programmes, and related efforts pertaining to replication and meta-analysis (Nosek & Errington, 2020; Simons, Shoda, & Lindsay, 2017). In this context, an important dimension of personality measurement from big data is *methodology*, because the same set of big data might be analysed with any of a wide range of machine learning algorithms (e.g. random forests, support vector machines, and elastic net regression). Extending Cronbach's term to include methodology, UTOS then becomes MUTOS (Becker & Aloe, 2008).

Another important parallel with the methodological traditions in psychological research comes with the big data Vs of veracity and value, which respectively tie into the well-established concepts of reliability and validity found in psychometrics, as well as the concept of utility found in industrial-organizational psychology. *Veracity* refers to the quality and accuracy of big data, which is often lower for *individual* data points that are seemingly related to personality (e.g. a single Facebook 'like'), but can dramatically increase when appropriately *aggregated* across the multiple personality-related situations and time points that big data offer. Note that this point parallels the historical discovery of psychometrically reliable traits via the aggregation of personality-relevant behaviours (Epstein, 1983). In the big data realm, one may aggregate big data by variables (e.g. grouping the columns of a data set via principal components analysis or factor analysis), by subjects (e.g. grouping the rows of a data set via *k*-means clustering or hierarchical agglomerative clustering), or by variables and subjects jointly, such that personality-based signals can be built up and contrasted against other systematic effects and errors found in a vast sea (or messy matrix) of big data.

Value is a big data term that often refers to validity and/or utility, for example, personality-relevant big data can hold value to the extent that it predicts meaningful outcomes in academic, organizational, and other developmental, health, and life domains (i.e. validity) and/or for the fact that those predictions usefully inform personal or institutional goals (i.e. utility). Thus, the value in big data personality research is often dependent on one's perspectives and goals. Organizations might define and quantify value and utility as a financial return on investment (Cascio, Boudreau, & Fink, 2019). But for researchers, big data findings may hold value to the extent that they can contribute to traditional ways of accumulating knowledge about effect size, predictions, and their generalization, as learned from patterns of relationship over time (Cronbach & Meehl, 1955), and via natural and formal interventions (Chester & Lasko, in press).

Technology

Many big data definitions also emphasize the specific and often-sophisticated hardware and software *technologies* that allow for the intensive real-time and large-scale collection and management of personality-relevant data. In addition to commercial efforts, open-source frameworks such as Hadoop and Spark have been developed for this purpose (for a survey of these and other popular big data frameworks, see Inoubli, Aridhi, Mezni, Maddouri, & Nguifo, 2018; also see mobileQ, https://mobileq.org/, Meers, Dejonckheere, Kalokerinos, Rummens, & Kuppens, 2019). Interactions in games, immersive virtual reality environments, and conversations with AI avatars are just a few examples of modern technologies that may provide big data on personality.

Certainly, modern digital technologies will affect the nature of personality-relevant big data and their subsequent analysis in many ways, both intended and unintended. For example, video surveillance of employees may provide reliable data and analysis pertaining to customer service, in terms of sociability, conscientiousness, conflict avoidance, and so on. However, much like the classic Hawthorne studies (Roethlisberger & Dickson, 1939), when employees know they are being continuously surveilled, the implementation of data-gathering and automated technologies in the workplace may yield unintended negative effects on employees responses/behaviours (e.g. approaching customers more than is necessary), and reactions will likely vary depending on personal characteristics (Yost, Behrend, Howardson, Darrow, & Jensen, 2019) and demographic characteristics (e.g. age, gender, race/ethnicity, and socio-economic status). If one seeks to understand the culture of modern work, education, and home settings involving big data, then some form of personality measurement must be an essential part of it.

This raises an important concern surrounding the *objectivity* of big data measures of personality, which has obviously been a longstanding concern with traditional personality measures. Psychometrically, traditional personality tests attempt to minimize variance in construct-irrelevant factors, meaning that test scores are as unaffected as possible by both systematic and random errors related to test administration, content, and test responding. By contrast, just as researchers hope to find personality-relevant variance in big data, there is also a massive suite of systematic and random irrelevancies that an algorithm can be trained on, leading to confounds we tend to call *algorithmic bias* (Barocas & Selbst, 2016). At best, these irrelevancies will add random errors to the measurement of personality using big data (the rough equivalent of lowering Cronbach's alpha). At worst, incorrect conclusions and decisions may result if such irrelevancies vary systematically with participant characteristics (e.g. demographics), individual differences (e.g. personality itself, cognitive ability, and video game experience), and situational influences (e.g. Internet bandwidth, computer equipment). To provide an example that raises these issues, previous research has found that slightly more women, African Americans, Hispanics, and younger applicants used mobile devices to complete high-stakes employment assessments than other demographic groups (Arthur, Doverspike, Muñoz, Taylor, & Carr, 2014), and it is expected that those devices generally will not impair the application process relative to other application methods. Although mobile technologies tend to show small or no subgroup differences for traditional personality measures (Arthur, Keiser, & Doverspike, 2018), one should be sensitive to potential subgroup differences in big data, personality-oriented and otherwise, as there are then potential fairness and legal

DOI: 10.1002/per

implications in terms of ensuring the data are based on the equal access and use of technology.

Overall, it cannot be assumed wholesale that big data are personality relevant or can reliably predict more established measures of personality constructs. Historically, traditional computerized and pencil-and-paper versions of personality inventories have been closely examined psychometrically in many ways to determine the strength of construct measurement and the nature and extent of systematic construct irrelevancies (e.g. because of measurement method, item-specific content; Arthur, Glaze, Villado, & Taylor, 2010; Meade, Michels, & Lautenschlager, 2007; Salgado & Moscoso, 2003). By contrast, most sources of big data reflect a nature or purpose *other* than personality measurement and are often used for prediction but with very little consideration for construct measurement. Big data would therefore benefit from psychometric modelling and methods (even if these models/methods depart from traditional methods) to help ensure that personality construct variance is being measured, while minimizing or avoiding deleterious contextual effects (e.g. distortions and biases). And likewise, more traditional and psychometrically well-developed personality measures can contribute to and be compared with more organic forms of big data.

Technology can influence more nuanced and less obvious characteristics of personality measurement with big data as well. For example, perhaps theory and thinking suggest that a particular measurement should be captured at certain intervals (e.g. every 10 seconds), but the data were collected by a device that recorded measurements at longer intervals (e.g. 30 seconds) or perhaps were event-driven instead (e.g. the data were captured whenever certain patterns of movement, physiology, physiognomy, or location are exhibited). Although these measurement considerations generally and typically exist in all research design, the point here is to appreciate how the choice and design of technology interacts with conceptual considerations, practical constraints, and possible analyses.

Algorithms

Regarding this last point, the digital technologies that yield big data may also influence what analytical techniques are appropriate (and inappropriate) through the types of the data that they produce. For example, when personality-relevant big data can be collected, integrated, and analysed simultaneously on a real-time or streaming basis from multiple technological sources (e.g. video, cell phone, and social media posts), the modelling of the data (and also the prior cleaning and preparation of the data) may necessarily become complex along multiple levels. For instance, imagine personality big data, collected over time, dynamically predicting outcomes for students (i.e. engagement, teamwork, and grades), teachers (i.e. giving timely and supportive feedback and providing accurate information), and institutions (i.e. overall student retention and graduation rates). This type of complex prediction problem could be modelled with any of hundreds of types of machine learning algorithms, or it could be modelled explicitly in a multilevel, structural, and dynamic/streaming framework (e.g. Ippel, Kaptein, & Vermunt, 2019). In other words, big data may be analysed in myriad ways, and these choices have important implications for how and even whether research questions about personality can be answered with the available data.

While keeping this last point in mind, we will now focus on machine learning, which people often speak of as if it were a monolithic entity, when as noted previously, it actually comprises a huge array of analytic techniques. Appreciating that no single machine learning algorithm can ever be considered uniformly superior across all contexts (Domingos, 2012), as well as appreciating the lack of guidance in personality research for selecting among machine learning algorithms, we can at least present how machine learning algorithms are organized into two broad and distinct categories: clustering and predictive modelling.

Clustering is similar in spirit to traditional methods of profile analysis, where cases (rows) can be usefully summarized by clusters that are internally consistent yet externally distinct. Popular clustering methods include *k*-means, hierarchical agglomerative clustering, and spectral clustering. Like profile analysis or factor analysis, clustering methods have no external criterion for informing each case's probability of cluster membership, which is why clustering is also called *unsupervised learning* in the big data literature.

In a sense, the central aim of clustering is to group cases or rows of a data set such that the number of clusters/groups maximizes the between-cluster versus within-cluster variance (in the spirit of ANOVA). Most clustering methods assign each case a probability of membership to each cluster, where each case 'belongs' to the cluster with the highest associated probability. In the personality context, clusters might reflect profiles of traits that are informative and provide information beyond each trait considered as a main effect. For some time, there has been a strong and longstanding debate that suggests that typological approaches (idiographic, or profile-based, cluster-based, or type-based approaches) to personality are generally not robust predictively or parsimonious conceptually (Donnellan & Robins, 2010; however, see the aforementioned Molenaar & Campbell, 2009). But perhaps machine learning algorithms applied to big data will change that, bringing new insights and additional value to the typological approach: for recent inspiration, see the four personality types identified using machine learning techniques applied to a Big Five measure in Gerlach, Farb, Revelle, and Amaral (2018) and an important rebuttal by Ones and Wiernik (2018).

Predictive modelling is similar to regression analysis, where, unlike clustering, a dependent variable (or variables) is predicted from a set of continuous and/or categorical predictors (often referred to as *features* in predictive modelling). Having a criterion is the hallmark of predictive modelling with *supervised learning* algorithms. As of this writing, the *caret* package within the R programming language contains 238 machine learning algorithms for prediction, a huge number that only keeps growing (see https://topepo.github.io/caret/available-models.html; Kuhn, 2019). Although the algorithms differ widely, they share the central goal of

DOI: 10.1002/per

achieving robust prediction by doing two things not typically found in traditional statistical analyses of personality data:

1 *Disturb the model*. Certain machine learning algorithms tend to combine predictive results across hundreds of weak models to result in a stronger prediction (e.g. averaging across hundreds of trees in a random forest). Or in a similar vein, other algorithms might examine a grid of possible model parameters (hyperparameters) to 'tune' the model in the search for complex yet robust relationships (e.g. the learning rate parameter in gradient-boosted machines; the cost parameter in support vector machines).
2 *Disturb the data*. Virtually all machine learning models are concerned with some form of cross-validation that keeps the data for model development and the data for prediction separate. In 10-fold *cross-validation*, for instance, the data set is randomly divided into 10 sections or folds; 9/10ths of the data is considered the *training set*, used to develop the model; and then predicted values generated for the *test set*, or the 1/10th of the data that did not participate in model development. This process is repeated nine more times so that every case gets to be in the test set with a predicted value. *Bootstrapping* methods contain another form of cross-validation, where a model is first estimated from the bootstrap sample (subsampling with replacement); then predictions are made for those cases from those data that did not participate in a given model; and this process is repeated a large number of times.

By disturbing both the model and the data, the wide range of machine learning techniques attempt to seek out the best model among those searched (#1 above) in terms of successful predictions (e.g. lowest mean-squared error) under cross-validation (#2 above).

In addition to clustering versus predictive modelling, machine learning algorithms can also be considered in terms of their *interpretability* (Ribeiro, Singh, & Guestrin, 2016a), which has obvious important ties to scientific understanding, policymaking, fairness, ethics, and efforts to improve both data and algorithms. Some algorithms are more interpretable than others. For example, it is relatively straightforward to explain the personality profiles of cluster means (centroids) in *k*-means clustering or *k*-nearest-neighbour prediction, or to explain the regression coefficients of personality predictors found in lasso or elastic net regression. In these cases, the models and processes by which these algorithms evaluate the associations between inputs and outputs are relatively transparent. By contrast, there are many *black-box* algorithms that, as the name suggests, are relatively opaque. Often the complexities explored by these algorithmic techniques do not have a closed-form function and are therefore much more difficult, if not impossible, to interpret (e.g. artificial neural networks tune network weights within arbitrary layers of hidden nodes; random forests average across hundreds of trees of variables; support vector machines use nonlinear profile matching along classification and prediction boundaries; see James, Witten, Hastie, & Tibshirani, 2013). Explanation versus prediction often (but not always) reflects a trade-off found in machine learning, where one must decide whether to gain some predictive power at the expense of a more straightforward and interpretable model (or vice versa).

Given this challenge of interpretability, why should one even consider using machine learning algorithms in personality measurement with big data? There are two very practical reasons worth emphasizing: (i) the number of variables exceeds the number of cases in a data set (e.g. because text, audio, and social media data sets are vast), meaning that traditional analyses such as multiple linear regression are impossible (i.e. the matrix will not be invertible; Fan & Li, 2006), and (ii) the researcher seeks to go beyond traditional analyses, to see whether robust complex relationships (ones that may not be specified a priori) can be located, and prediction can improved without overfitting the data (i.e. mistaking sampling error variance for actual complexity; Hawkins, 2004).

To move beyond these general considerations and learn more about what machine learning is and how to apply it, most personality psychologists would benefit from the practical guide to big data in psychology by Chen and Wojcik (2016), as well as the psychology-oriented big data tutorials pertaining to text mining or web scraping (e.g. Landers, Brusso, Cavanaugh, & Collmus, 2016), data processing and predicting outcomes (e.g. Kosinski, Wang, Lakkaraju, & Leskovec, 2016), and meta-analysis (e.g. Cheung & Jak, 2016).

Impact

The last major theme contained in many definitions of big data reflects the *impact* of big data, or how the output from big data algorithms is used as feedback for our behaviours: informing, interacting with, and potentially magnifying, mitigating, or otherwise influencing them. These reciprocal effects between algorithms and behaviours might only be understood within short time spans, although it is possible that habits, personality traits, and behaviours could be affected in the longer term as well. This type of iteration and impact should be considered in the context of the range of settings in which big data are used, from the settings that might be relevant at the individual level, to the broad and aggregated settings of big data analysis that might be relevant at the community and policy levels. For example, if we consider how a user's personality affects how and when they follow up on recommendations from a smartphone health app, then the individual-level impact on health, privacy, and safety over time may be of paramount importance. However, for scholars and policy makers who are focused on how personality data from social media and résumé websites are used to recruit and hire employees, then the societal-level focus may be on trends in national productivity, worker rights, or legal liabilities. Applications in all these settings, and at all these levels, have ethical challenges for applying knowledge concerning personality, where enumerating, discussing, and weighing the pros and cons often do not lead to obvious or single solutions and therefore must be explored deeply and on a continuous basis.

DOI: 10.1002/per

The potential impact of big data that is perhaps most relevant to personality researchers is how these data might increase the scientific understanding of personality. For example, large behavioural data streams may identify new indicators and aspects of personality that have not previously been revealed via traditional personality measures, such as Big Five measures based in the lexical tradition (Bleidorn, Hopwood, & Wright, 2017). Moreover, novel types of behavioural variables found in digital trace data may advance personality research through revealing new insights about trait content and trait-related processes. Big data insights such as these can lead to developing measures of new or expanded constructs in the future, ones that may amplify the original big data signal over the noise even further. For example, the lexical hypothesis, upon which many Big Five measures of personality have been developed, assumes that all personality can be captured in our language (e.g. introspective or retrospective self-report questionnaires). However, if some aspects of personality are better captured through behaviour, then perhaps existing personality constructs will be refined further, or new personality constructs will emerge from vast streams of personality-relevant behavioural data. Also, given the intensive sampling that big data offer, new patterns of personality change may be discovered that are reliable and predictable, if not dispositional.

Other aspects of big data may shed light on personality development and the expression of personality in various contexts. The ubiquity of big data, the large diverse samples they offer, and their temporal granularity promise to yield greater insight into the reliability of personality indicators over one's lifespan and across cultures (see Bleidorn et al., 2017). Additionally, big data and machine learning might complement current advances in personality research, for example, through measuring and analysing massive amounts of data collected across the lifespan of huge samples that traditional technologies and statistical methods simply cannot. Through the richness of big data, we might gain more ecologically valid insights into personality development (and stability) across the lifespan, and how personality facilitates, limits, or otherwise interacts with situational factors. For big data that are multicultural and international in nature, this big data approach promises to help researchers gain a deeper understanding of the culture-free (etic) and culture-specific (emic) components of personality, and not merely enhance predictive validity (although these are often interrelated goals). As a prime example, the large, culturally diverse data sets offered by the many digital communication platforms used throughout the world provide a wealth of personality data where researchers can compare the quality, reliability, and validity of big data indicators of personality in many different populations, in many different settings, and over long periods of time.

Applying this framework to big data personality measurement

As our discussion of this framework suggests, the measurement of personality using big data is complex and involves multiple interrelated categories: information, technology, algorithms, and impact. Thus, the reader should appreciate that the value of personality measurement using big data depends on the *combination* of these categories—none of these should be considered in isolation. For example, the technology used to collect data (e.g. a social media platform) will determine the types of personality relevant big data available (e.g. large and sparse data sets of social media 'likes' where there are more variables than cases). Consequently, these types of data may be more readily analysed by a certain algorithm or set of algorithms (e.g. using clustering algorithms to reduce the dimensionality of the data, and then predictive algorithms to classify the media platform's users based on some dimension of personality). In turn, the algorithmic approach chosen may be more or less interpretable than other approaches, having implications for the impact of using machine learning algorithms in situations where a clear understanding of what has driven a prediction is useful or necessary, such as when the use of big data may or may not be legal (e.g. employee recruitment) or safe (e.g. medical treatment recommendations). A good practice when planning big data personality analyses is to generate options within these four categories, appreciating how the choices made in one category often influence the options available in another (e.g. the choice of technology used to collect big data influences the information available). Mapping out multiple scenarios in this way, and comparing them with one another, can also be helpful in this planning process.

BIG DATA CHALLENGES TO RELIABILITY, VALIDITY, FAIRNESS, AND PRIVACY

Big data involving personality-relevant information, accompanied by robust predictions, have the potential to improve our understanding of personality. In fact, the reason researchers and practitioners are investing in big data and machine learning is to obtain new and better personality insights, obtained through more sophisticated-yet-robust relationships, and a better overall understanding of personality constructs. However, as exciting as this might sound, benefits do not accrue automatically by bringing big data and machine learning together. Instead, in line with the hypothetico-deductive cycle, personality researchers must continue to iterate between top-down intentional forms of big data (those resulting from measures based in personality theory) that inform and are informed by bottom-up incidental big data (those that happen to be available). Such an iterative process of empirical analysis motivating the refinement and revision of psychological theory (and vice versa) has a long scientific history (Fiedler, 2018). This process will continue with big data analyses as it has with traditional data analyses, where, both within and across research endeavours, we continuously accrue information about big data reliability and validity that informs personality theory.

Reliability

The nature of big data poses many challenges to quantifying sources of construct-relevant variance and error variance,

DOI: 10.1002/per

and how to create an appropriate reliability model in doing so. Using the framework, which we previously introduced, it is likely that appropriate reliability estimates will be heavily influenced by the unique combination of information, technology, and algorithms within each study or application. The complexity and messiness of big data—particularly when the number of variables exceeds the sample size—will inevitably demand novel approaches to estimating reliability. Perhaps in some cases, traditional concepts of reliability can still be of use (e.g. internal consistency, test–retest, parallel forms, and interrater agreement), but we suggest that new conceptions of reliability, or hybrid estimates that combine traditional measures of personality with big data, might find new ground and new importance in the context of big data.

For example, a big data set may contain movie ratings, but the matrix is sparse because individual subjects view only a very small subset of those movies. Therefore, reliability and agreement are challenging to estimate (e.g. rank-order stability, absolute accuracy, and factor structure). To add even more complexity to this modelling problem, ratings might be time dependent (e.g. holiday movies are rated higher over the holiday of interest); person dependent (e.g. only certain people celebrate certain holidays); and people drop in and out of the data set arbitrarily. These and other factors may have independent and joint effects on big data as a reflection of personality. Here, various forms of network analyses (Epskamp, Rhemtulla, & Borsboom, 2017) may take the place of traditional reliability estimates (see Christensen, Golino, & Silvia, 2020, and Costantini et al., 2019, for inspiration specific to personality research).

Thus far, we have only discussed the reliability of personality big data in relation to the characteristics of the data themselves, but the technology used to collect the data also carries important implications for reliability. Returning to our previous example, not all movie raters in our hypothetical big data set are watching movies on the same device or in the same environment. Some subjects in the data set may have rated a movie after watching it on a 60-inch smart TV with a fast fibre optic Internet connection in their spacious home theatre, whereas others watched the same movie on the 5-inch screen of their iPhone with a sporadic Internet connection on the subway train during their commute home from work. This begs the empirical question, how reliable are such ratings across different devices and different settings? At this point, we cannot answer this question in too much detail. However, it seems clear that understanding systematic variation in the reliability of big data indicators of personality across such technological contexts would serve a very practical purpose that informs appropriate data collection and interpretation. For example, time-intensive data-collection technologies, along with the types of data that are collected, might affect the intervals at which we should assess (and re-assess) the reliability of big data personality measures. Importantly, the interest in personality-related big data opens the door to understanding real-time changes in data quality, reliability, validity, and subgroup differences—and how they are related.

We have discussed just a few of the implications the information and technology aspects of big data may have for reliability, but the complexity of the analyses allowed for by algorithms themselves also pose challenges to understanding the reliability of personality measurement with big data. In a recent example, machine learning algorithms have proven useful in attempts to analyse idiographic or within-person models of personality, meaning that each person contains variation and covariation in their own trait expressions that can be modelled over time. Beck and Jackson (2019b) used experience sampling method (ESM; Csikszentmihalyi & Larson, 2014) data to model the time-lagged and contemporaneous relationships between within-person manifestations of personality using the interpretable machine learning method of the graphical lasso (Friedman, Hastie, & Tibshirani, 2008; an extension of Tibshirani, 1996). The authors found contemporary idiographic models to be relatively consistent over time, whereas lagged models were not; however, both models were found to demonstrate individual differences in consistency over time. Specifically, the results indicated that within-person networks of personality variables were relatively consistent with those measured in a second wave ESM study conducted a year later when the personality variables were measured at the same time points (i.e. contemporaneous). However, when ESM measurements were lagged by about 4 hours (to model the daily dynamics of personality variables), the models were not consistent with those measured in the second wave of the study. Although this represents a very novel research application of big data, it is interesting to consider how machine learning models might begin to shed light on scientific questions that involve the prediction of dynamic within-person personality processes unfolding over time and individual differences in their reliability.

To move research of this nature forward even further, others have argued for understanding the psychometric properties of experience sampling personality state measures (Beckmann & Wood, 2017), an important application of the need for personality researchers to improve their understanding of the reliability of big data. Although much research has been conducted on big data indicators of personality (see the meta-analysis by Azucar et al., 2018, for a list of studies predicting Big Five personality traits from digital footprints on social media), most of this research has been focused on the prediction of traditional measures of personality (Wright, 2014), such that outside of this sphere, much about the reliability and other psychometric properties of big data indicators of personality have yet to be evaluated. Reliability can take many forms that are analogous to traditional ones, even if it is modelled differently, for example, internal consistency, longitudinal, and alternate forms. Psychometric modelling of big data will be an important direction, because reliability is an important feature of all data, yet as we have already emphasized, personality-relevant big data are often messier (less structured, more variables) than the personality data and items coming from traditional psychological tests.

In short, reliability is just as important for understanding personality in the big data era as it has always been, because there are scientific, practical, and ethical imperatives to

DOI: 10.1002/per

separating variance relevant to constructs (e.g. personality, motivation, knowledge, and otherwise) from variance irrelevant to constructs (e.g. demographics, rater biases, item-specific variance, and otherwise). We thus look forward to seeing personality psychologists informing data scientists (e.g. applied statisticians and computer scientists), and vice versa, when developing new approaches to evaluating the reliability of personality-relevant variance within big data sets.

Novel measurement approaches

In traditional measurement, the purpose of reliability as a part of construct validation has been to provide evidence that the observed variance in scores on a measure of personality is related to the underlying latent personality construct being measured (Borsboom, Mellenbergh, & Van Heerden, 2004). This task can be extremely difficult for big data, because researchers and respondents have less control over how the data are produced. For example, the very same technologies and predictive algorithms that collect big data in the first place (e.g. Facebook likes, Twitter posts, Netflix movie ratings, and Amazon product ratings) are being used to select and tailor interventions that inform, amplify, reduce, or otherwise condition behaviour (e.g. future consumer purchases). Turning back to our movie example, a person's movie choices and ratings lead to movie recommendations—and an algorithm's recommendations are based on the continuous analysis of viewer choices and ratings. The general point here is whether algorithmic recommendations based on personality-related behaviour are premature, just right, too late, or something else. In this way, personality is in a network of dynamic relationships with behaviour and the interventions received. This design is by intent; however, it makes the behavioural big data harder to interpret as purely trait based (Kosinski, Matz, Gosling, Popov, & Stillwell, 2015).

Therefore, anchoring our predictive modelling decisions in theory and remaining mindful of possible confounds, while examining the content validity of big data personality indicators of personality, is our first line of defence against training our models on such construct irrelevancies. Certainly, well-trained expert raters can attempt theoretical justification for the inclusion or exclusion of potential big data personality indicators, but they are inevitably limited to the potential indicators present in the data set, which might contain both construct irrelevancies and construct deficiencies. Coupled with the rational/conceptual approach of expert rating is the empirical approach of applying the nomological network (Cronbach & Meehl, 1955), meaning that perhaps we can 'back into' personality constructs by demonstrating sensible big data patterns of convergent and discriminant validity. Of course, such work is easier said than done, as all this presumes that machine learning algorithms and data at least lean towards being interpretable. Perhaps that leaning will become even stronger as data scientists understand that construct validity remains very important (Bleidorn & Hopwood, 2019; Tay, Woo, Hickman, & Saef, 2020)—perhaps even more important than ever before given the nature of the data.

Of course, up until this point, our discussion of construct validation has focused on identifying big data indicators that reflect the Big Five and other established personality constructs. Beyond this application, however, there lies the idea, the potential, for advancing personality theory by introducing *novel unique personality predictors* that may reflect, extend, or even replace existing theories. For instance, conscientiousness is viewed as a personality trait, yet perhaps new technologies for measuring conscientiousness help to redefine the construct by measuring how it changes in its nature over time and how it predicts behaviour over time (e.g. as a process in school, Corker, Oswald, & Donnellan, 2012). On the other hand, there is also the strong caution and potential danger for complex predictive algorithms applied to messy big data sets to contribute to the *jingle-jangle personality constructs* (Kelley, 1927; Thorndike, 1904), essentially by empirically repackaging and combining data (new wine) that may reflect established personality constructs (old bottles). In considering these promises and perils for novel construct definitions and measurement, note that the massiveness of big data combined with the opacity of machine learning could lead to *innovative and personality-relevant variance going unnoticed* if we are not careful. To avoid such confusion, it will be illuminating for research to continue to compare and contrast personality constructs (established versus novel), personality measures (traditional Likert scales versus adaptive item-response-theory-based tests versus big-data-based), and analysis tools [exploratory factor analysis versus confirmatory factor analysis (CFA) for clustering; linear regression versus structural equation modeling (SEM) versus machinelearning algorithms for prediction].

The practical and scientific value of big data in personality measurement

For many, the most exciting aspect of big data in personality measurement lies in its many real-world applications. Interactive technologies allow for targeted intervention on human emotion, thought, and behaviour, and this dynamic data, in turn, can be related to personality, where a large body of research has demonstrated relationships between personality and health and well-being (Strickhouser, Zell, & Krizan, 2017). Research has also found relationships between personality and compliance to medical treatment (Umaki, Umaki, & Cobb, 2012) as well as prescription drug regimen adherence (Axelsson, Brink, Lundgren, & Lötvall, 2011; Christensen & Smith, 1995). Thus, machine learning algorithms may offer exciting opportunities to improve human health and well-being through personalized interventions, using big data to detect and influence personality, with the goal of forecasting and improving real-world health outcomes (e.g. within health risk evaluation, clinical treatment, and public health programmes; Chapman, Hampson, & Clarkin, 2014). In fact, a recent article demonstrated that adding personality measures to a machine learning medical risk model predicting log hazard rates of mild cognitive impairment significantly improved the model's overall prediction measured by Nagelkerke's

DOI: 10.1002/per

pseudo R^2 and discrimination as measured by area under the receiver operating characteristic curve (Chapman, Lin, Roy, Benedict, & Lyness, 2019).

Robust out-of-sample prediction of potentially complex relationships is obviously a major strength of machine learning and big data, as we have noted; yet we should not always rely on those 'black box' predictions alone, because we usually want to know *why* prediction happens and not just *that* it happens. Thus, the trade-offs between interpretability and prediction must be understood and managed when deciding among machine learning and traditional analysis options. Fortunately, however, some recent attention in machine learning has turned to interpreting these black box models (Ribeiro et al., 2016a). Although there is debate about the appropriateness of the interpretation of such models, especially for high-stakes applications (Rudin, 2019), if such model interpretation can be achieved, it may be useful when attempting to apply psychometric evaluations of validity to these analytical techniques, as knowing what variables are driving predictions is essential to substantive understanding and discovery with big data personality measures. Although still in their infancy, machine learning interpretation techniques may hold promise for psychology researchers who wish to draw sound scientific inferences about personality from big data. One such tool, local interpretable model-agnostic explanations (Ribeiro, Singh, & Guestrin, 2016b), explains any black box model by approximating the predictions of the original model with a locally interpretable one (a local surrogate model) trained on perturbations of the original data (permuting or removing predictors in the model). Interpretability then can better emerge from the local region or regions of interest, a useful approach when personality researchers use more complex forms of machine learning.

Although we have mentioned some of the obvious ethical and privacy concerns surrounding big data personality measurement, what if people at risk of obesity, as identified by their personality (Sutin, Ferrucci, Zonderman, & Terracciano, 2011), could be detected and assisted by big data technologies (e.g. healthy food choice or exercise recommendations)? To the extent that personality is malleable, could interactive technologies influence the development or expression of personality traits or facets that contribute to successful weight loss in those seeking to lose weight (Soini, Mustajoki, Eriksson, & Lahti, 2018)? This example and others like it are not just academically interesting; they are important because healthcare outside the hospital involves *self-management*, a critical component of many technologies that inherently will be influenced by personality. By understanding personality, these health-oriented technologies can track human wellness and disease even more effectively (Erdmier, Hatcher, & Lee, 2016; Jeong, Bychkov, & Searson, 2019), and they can be tailored to be more responsive and useful. Of course, there are also many applications outside of the healthcare industry where personality-based interventions might improve people's quality of life at both the individual and societal levels. Indeed, one can imagine that the detection and/or manipulation of personality could be usefully exploited to create personalized interventions that improve many important outcomes, while realizing that more empirical work investigating this hope is desperately needed.

Fairness

Fairness issues are critical to consider when capturing personality data, applying machine learning, and then reporting and acting on analytic results. Although Internet access penetrates the majority of the US population and has been described as a 'basic utility for social inclusion' in technology-centric societies (Van Deursen & Van Dijk, 2019, p. 355), the type, speed, and consistency of Internet access in the world, or even in specific neighbourhoods, can certainly be uneven (Marler, 2018). In fact, decades of attention have been paid to the 'digital divide', or the general disparities in access to, usage of, and proficiency in Internet and communication technologies across many key demographics such as age (Van Volkom, Stapley, & Amaturo, 2014), gender (Ching, Basham, & Jang, 2005), race (Jackson et al., 2008), and socio-economic status (Huffman, 2018), as well as within many critical domains of human activity and endeavour, such as education (Rowsell, Morrell, & Alvermann, 2017), employment (Lindsay, 2005), and healthcare (Mackert, Mabry-Flynn, Champlin, Donovan, & Pounders, 2016). These differences in access to Internet and communication technologies serve as an important reminder for personality researchers to be sensitive to the representativeness of big data. Representativeness here means that using technology to understand personality in diverse populations requires sampling in a similarly diverse and representative manner, moving beyond populations that are westernized, educated, industrialized, rich, and democratic (WEIRD; Henrich, Heine, & Norenzayan, 2010).

Consequently, being fair in the era of big data begins with a sensitivity to critical differences in Internet accessibility and technical proficiency, as well as other regional, linguistic, and race/ethnicity demographics of concern to the researcher. These factors, and more, can influence how (or even whether) personality inferences can be made appropriately from big data. The key fairness question to be investigated here is as follows: can personality be inferred in a similar way across subgroups of interest? If personality constructs differ in essential ways across subgroups, then many fairness problems arise, for example, it would be challenging or even unfair to compare subgroups directly with one another, and criterion-related validity would mean different things for different subgroups.

Traditionally, this issue has been of concern, where psychometric judgements of construct similarity involve examining factor structure and measurement invariance across subgroups of interest (Chiorri, Marsh, Ubbiali, & Donati, 2016; Marsh et al., 2010). This traditional approach could be usefully extended to personality-relevant big data. In traditional measurement, psychometricians will often first undertake conceptual steps (e.g. reviewing item content for construct relevance and cultural sensitivity) and empirical steps (e.g. investigating measurement invariance and

DOI: 10.1002/per

predictive invariance), to determine whether personality measures tend to capture traits reliably and exhibit empirical relationships in the same way across subgroups (Millsap, 2011). Even though the goals of measurement and predictive invariance cannot be simultaneously fulfilled in the strictest mathematical sense (Millsap, 2007), it usually remains practically informative to determine how closely these goals are met.

To the extent there is strong evidence against measurement invariance, for instance, then mean differences between subgroups become more driven by the quirks of the items than by the construct. And with evidence against predictive invariance, an overall regression line will make systematic overprediction and underprediction for the target subgroups of interest that should also be considered. Granted, the approach and application of invariance analyses will look different when based on big data and related algorithms than when based on traditional personality measures (e.g. applying some form of within-and-between group network analysis to personality items instead of a CFA). But the point is that without measurement invariance analyses of some sort, construct-level comparisons between subgroups on personality-relevant big data are compromised, as well as a deeper understanding about the substantive nature of algorithmic bias (construct irrelevance) in big data. But if we do not even know what or how much variance observed is construct relevant (per the discussion earlier on reliability and validity), then we cannot begin to investigate measurement and predictive invariance.

Privacy

Last but hardly least, big data pose ethical challenges involving those who use the data to infer personality, as well as the people from whom the data are collected. Privacy concerns have been longstanding in the health sciences (e.g. Murdoch & Detsky, 2013) and are emerging in the individual privacy domains (e.g. General Data Protection Regulation in Europe; Family Educational Rights and Privacy Act, Fair Credit Reporting Act, and state laws in the US), but these concerns are relatively new to personality research (although for guidance, see Kosinski et al., 2015). Ethical sensitivity should certainly accompany all psychological big data research (Mansour, 2016), especially considering that only a small subset of characteristics may be necessary to re-identify individuals within an anonymized data set (Rocher, Hendrickx, & de Montjoye, 2019).

Specific to personality research, note that the presence and type of big data technologies that are used will likely change people's behavior meaningfully, as the perceptions and reality of compromised privacy and anonymity with regard to big data continue to grow, and as more diverse data and complex analytic tools come together more often, more quickly, and on a broader scale. As people become aware that their data at home, at work, at school, and online are being combined and mined, they may wish to share their data when the perceived outcome is positive (e.g. better house purchase and better medical advice), and wish to keep their data confidential when the outcome is likely to be neutral or negative (e.g. when data and algorithmic predictions are being sold to the highest bidder or when negative information might undermine potential job prospects). Or, being inured to big data privacy issues, they might not change their behaviour at all under the assumption that the privacy of their data always runs the risk of compromise.

MEETING THE CHALLENGES OF BIG DATA AND MACHINE LEARNING IN PERSONALITY RESEARCH

Researchers outside of the field of psychology have experience using a wide variety of data analytic techniques for modelling complex systems that might prove valuable in measuring personality with big data (see Gerlach et al., 2018 for an example of a collaboration between engineers, psychologists, and physicists to identify personality types using big data). Overall, it seems clear that personality psychologists are absolutely essential to the work of data scientists, if they are not data scientists themselves, because big data and machine learning algorithms critically depend on what traditional research depends on: framing the personality modelling problem appropriately (e.g. clustering versus prediction) and selecting construct-relevant and reliable measures and outcomes (e.g. personality and other psychological characteristics; Flake & Fried, 2019). Only then can the research and results be not only analytically sound but also context-appropriate and informative towards meeting personality research and practice goals (Lodge, Alhadad, Lewis, & Gašević, 2017).

Fortunately, personality researchers bring valuable expertise and training that reflect a unique set of skills that we believe will prove of increasing importance towards interpretable, practical, ethical, and defensible big data analyses: e.g. an understanding of personality constructs, the psychometrics training to develop and evaluate measures of those constructs, and study design and ethics training relevant to the context of social sciences research involving human subjects. With its long history of concern for ethical principles such as privacy, confidentiality, and informed consent, psychologists in general are particularly well-suited to develop big data personality assessments and to assist in the development of guidelines and laws that help to determine the ethical applications of big data. Not only have psychometric and ethical concerns been a driving force for many psychological researchers who study topics such as ethics in psychological testing, diversity and inclusion, and adverse impact; psychologists have also helped to codify the findings of these bodies of research into practical guidelines, as they have with the development of the *Standards for Educational and Psychological Testing* [American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME), 2014], the *Uniform Guidelines on Employee Selection Procedures* (Equal Employment Opportunity Commission, 1978), and the *Principles for the Validation and Use of Personnel Selection Procedures* (Society for Industrial

DOI: 10.1002/per

and Organizational Psychology, 2019). Taken together, personality researchers are in a unique position to understand, learn from—and teach others—the important and hard lessons involving the balance or trade-off between the psychological informativeness of big data personality approaches and the ethical challenges when using such personality-relevant technologies.

However, it is important for a personality researcher to understand that when sharing these traditions while engaging with big data and machine learning, their discipline often will join many others at the table that have their own expertise, goals, and unique priorities (e.g. data science, applied statistics, healthcare, economics, political science, and human resources management). Thus, a multidisciplinary approach is both challenging and heartening, because more substantive conversations and context around big data and machine learning algorithms can be had, giving greater meaning and closer criticism around the prediction-versus-explanation tradeoffs in machine learning (Yarkoni & Westfall, 2017). For personality researchers, prediction and explanation are part of an iterative process designed to deepen our understanding of the psychological nature of personality and the corresponding habits of human thought, attitudes, and behaviour.

FUTURE DIRECTIONS FOR BIG DATA PERSONALITY RESEARCH

As we enter this exciting new age of big data, we encourage personality researchers to reflect on 14 important aspects of big data personality analyses, to create more synergies (and fewer competitions) between prediction and explanation in the personality domain.

Information: Understanding how to approach big data

1. Collaborating with multidisciplinary colleagues in quantitative disciplines (e.g. computer scientists, data scientists, applied statisticians, and other social scientists), who have expertise and experience in contending with big data having multiple formats, along with messiness and missingness, and applying algorithms for clustering and dimensionality reduction.
2. Researching the reliability of big data collections of features (e.g. extracted social media themes) across time, samples, and settings, including collaborating to develop appropriate theory and statistical models that estimate those reliabilities in big data contexts.
3. Establishing the construct relevance of big data personality indicators being used via machine learning analyses (e.g. interpretation of clusters in *k*-means clustering or neural networks) as it can be compared with traditional reliability analysis (e.g. interpretation of factors and factor loadings in a CFA).
4. Striving for substantive interpretability of personality data; for example, develop theory and conduct empirical studies aimed at exploring and understanding the content validity of big data personality predictors.

Technology: Understanding the digital technologies that give rise to big data

5. Collaborating with multidisciplinary colleagues who have expertise and experience in the implications of using different digital devices, infrastructures, and software to collect, manage, store, or otherwise curate big data collections.
6. Researching the reliability of big data collections of features across different technology platforms from which the same data might be collected (e.g. laptop versus cell phone; different operating systems).

Algorithm: Adopting and evaluating algorithmic approaches to personality measurement

7. Collaborating with multidisciplinary colleagues who can inform personality research, who have expertise and experience in selecting and justifying the analytical approach chosen to analyse data, for example, deciding between different prediction models, training and tuning those models, and accurately interpreting their results.
8. Emphasizing analytic procedures that avoid overfitting models to personality data (e.g. *k*-fold cross-validation and bootstrapping), which is an emphasis in machine learning, but can be applied equally usefully in traditional modelling.
9. Investigating predictive patterns involving personality measures: statically, in terms of convergent, discriminant, and criterion-related validity; and dynamically, where mediational and multilevel (cross-level) relationships can be tested with longitudinal big data.
10. Evaluating the nature of big data collected over time—as well as results from the clustering and prediction algorithms applied dynamically to those data over time—such that inferences about both populations and personality can be drawn from big data, and the scope of the generalizability and malleability of personality can be further broadened, understood, and advanced.

Impact: Fairness and ethics

11. Striving to ensure that personality measurement is fair across demographic subgroups (e.g. race/ethnicity, gender, and culture), while realizing that fairness is a concept that encompasses broad issues such as cultural sensitivity, conflicting definitions such as equity versus merit, and equal opportunities to provide data.
12. Detecting and reducing personality-related biases and other irrelevancies detected by algorithms. Bias is a statistical concept, referring to empirically reliable subgroup differences in personality that are due to construct irrelevancies in the data, the models, or their combination.
13. Ensuring that personality data collection (e.g. data privacy, anonymity, and security) and data use (e.g. analysis and interpretation) are sensitive to and consistent with updated professional, legal, and ethical standards (e.g. AERA, APA, NCME, 2014; Equal Employment Opportunity Commission, 1978).

DOI: 10.1002/per

14. Transparently reporting the process of personality data collection and analysis, disclosing and reflecting on any key limitations alongside any key benefits. Open science practices can critically assist in improving transparency (e.g. preregistration, sharing relevant measures and protocols, and sharing the variable codebook and code for analyses, if not the data set itself).

CONCLUSION

This paper has outlined some of the promises and challenges for personality researchers interested in developing and collecting big data from new digital technologies, coupled with the machine learning algorithms applied to big data (and most any psychological data). For over a century, psychological researchers have accumulated the wisdom from hard-won lessons concerning the measure development, analysis, and evaluation process, for example, generating item content; refining measures psychometrically; examining correlational patterns with other measures, manipulations, and real-world outcomes. This wisdom seems hardly outdated in the big data era—perhaps it is even more important—and much of it is found in codified professional standards of measurement, as found in the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 2014), now in its third edition. Personality researchers can judge for themselves, as they are engaged in collecting personality-related big data with new (intensive, unobtrusive) data-collection technologies and sophisticated clustering and predictive algorithms, whether they will likely benefit from revisiting and making use of our psychometric concepts and tools.

Concluding with an even bigger picture, we would argue that big data not only has much to offer to personality psychology, but also vice versa, meaning that it is critical for the expertise, accomplishments, and history of personality psychology to shape how the big data and analytics community can more effectively and efficiently forge useful pathways towards achieving new and rapid insights about personality (habits of human attitudes, thoughts, and behaviour). To that end, personality psychologists must continue to build and play a part in multidisciplinary communities of interest focused on future developments, applications, and evaluations of big data and machine learning. Trends seem to be moving in this direction—perhaps we could apply machine learning algorithms to big data personality researchers themselves to accelerate that trend.

REFERENCES

Adjerid, I., & Kelley, K. (2018). Big data in psychology: A framework for research advancement. *American Psychologist*, *73*, 899–917. https://doi.org/10.1037/amp0000190

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Aroganam, G., Manivannan, N., & Harrison, D. (2019). Review on wearable technology sensors used in consumer sport applications. *Sensors*, *19*, 1–26. https://doi.org/10.3390/s19091983

Arthur, W., Doverspike, D., Muñoz, G. J., Taylor, J. E., & Carr, A. E. (2014). The use of mobile devices in high-stakes remotely delivered assessments and testing. *International Journal of Selection and Assessment*, *22*, 113–123. https://doi.org/10.1111/ijsa.12062

Arthur, W., Glaze, R. M., Villado, A. J., & Taylor, J. E. (2010). The magnitude and extent of cheating and response distortion effects on unproctored internet-based tests of cognitive ability and personality. *International Journal of Selection and Assessment*, *18*, 1–16. https://doi.org/10.1111/j.1468-2389.2010.00476.x

Arthur, W., Keiser, N. L., & Doverspike, D. (2018). An information-processing-based conceptual framework of the effects of unproctored internet-based testing devices on scores on employment-related assessments and tests. *Human Performance*, *31*, 1–32. https://doi.org/10.1080/08959285.2017.1403441

Axelsson, M., Brink, E., Lundgren, J., & Lötvall, J. (2011). The influence of personality traits on reported adherence to medication in individuals with chronic disease: An epidemiological study in West Sweden. *PLoS ONE*, *6*, 1–7. https://doi.org/10.1371/journal.pone.0018241

Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences*, *124*, 150–159. https://doi.org/10.1016/j.paid.2017.12.018

Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., & Stillwell, D. (2012). *Personality and patterns of Facebook usage*. *Proceedings of the 4th Annual ACM Web Science Conference*, 24–32. https://doi.org/10.1145/2380718.2380722

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *SSRN Electronic Journal*, *104*, 671–732. https://doi.org/10.2139/ssrn.2477899

Beck, E. D., & Jackson, J. J. (2019a). Within-person variability. *PsyArXiv. Advance online publication.*. https://doi.org/10.31234/osf.io/kavbp

Beck, E. D., & Jackson, J. J. (2019b). Consistency and change in idiographic personality: A longitudinal ESM network study. *Journal of Personality and Social Psychology*. Advance online publication, *118*, 1080–1100. https://doi.org/10.1037/pspp0000249

Becker, B. J., & Aloe A. M. (2008). *A framework for generalization in meta-analysis: Medical and social science examples [invited presentation]*. The 16th Merck-Temple conference on biostatistics, Philadelphia, PA.

Beckmann, N., & Wood, R. E. (2017). Editorial: Dynamic personality science. Integrating between-person stability and within-person change. *Frontiers in Psychology*, *8*, 1–7. https://doi.org/10.3389/fpsyg.2017.01486

Bleidorn, W., & Hopwood, C. J. (2019). Using machine learning to advance personality assessment and theory. *Personality and Social Psychology Review*, *23*, 190–203. https://doi.org/10.1177/1088868318772990

Bleidorn, W., Hopwood, C. J., & Lucas, R. E. (2018). Life events and personality trait change. *Journal of Personality*, *86*, 83–96. https://doi.org/10.1111/jopy.12286

Bleidorn, W., Hopwood, C. J., & Wright, A. G. (2017). Using big data to advance personality theory. *Current Opinion in Behavioral Sciences*, *18*, 79–82. https://doi.org/10.1016/j.cobeha.2017.08.004

Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*, 1061–1071. https://doi.org/10.1037/0033-295X.111.4.1061

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105. https://doi.org/10.1037/h0046016

Cascio, W., Boudreau, J., & Fink, A. (2019). *Investing in people: Financial impact of human resource initiatives* (3rd ed.). Alexandria, VA: Society for Human Resource Management.

DOI: 10.1002/per

Cattell, R. B. (1946). Personality structure and measurement. I. The operational determination of trait unities. *British Journal of Psychology*, *36*, 88–103. https://doi.org/10.1111/j.2044-8295.1946.tb01110.x

Chapman, B. P., Hampson, S., & Clarkin, J. (2014). Personality-informed interventions for healthy aging: Conclusions from a National Institute on Aging work group. *Developmental Psychology*, *50*, 1426–1441. https://doi.org/10.1037/a0034135

Chapman, B. P., Lin, F., Roy, S., Benedict, R. H. B., & Lyness, J. M. (2019). Health risk prediction models incorporating personality data: Motivation, challenges, and illustration. *Personality Disorders, Theory, Research, and Treatment*, *10*, 46–58. https://doi.org/10.1037/per0000300

Chen, E. E., & Wojcik, S. P. (2016). A practical guide to big data research in psychology. *Psychological Methods*, *21*, 458–474. https://doi.org/10.1037/met0000111

Chester, D. S., & Lasko, E. N. (in press). Construct validation of experimental manipulations in social psychology: Current practices and recommendations for the future. *Perspectives on Psychological Science*.

Cheung, M. W.-L., & Jak, S. (2016). Analyzing big data in psychology: A split/analyze/meta-analyze approach. *Frontiers in Psychology*, *7*, 1–13. https://doi.org/10.3389/fpsyg.2016.00738

Ching, C. C., Basham, J. D., & Jang, E. (2005). The legacy of the digital divide: Gender, socioeconomic status, and early exposure as predictors of full-spectrum technology use among young adults. *Urban Education*, *40*, 394–411. https://doi.org/10.1177/0042085905276389

Chiorri, C., Marsh, H. W., Ubbiali, A., & Donati, D. (2016). Testing the factor structure and measurement invariance across gender of the Big Five Inventory through exploratory structural equation modeling. *Journal of Personality Assessment*, *98*, 88–99. https://doi.org/10.1080/00223891.2015.1035381

Chittaranjan, G., Blom, J., & Gatica-Perez, D. (2013). Mining large-scale smartphone data for personality studies. *Personal and Ubiquitous Computing*, *17*, 433–450. https://doi.org/10.1007/s00779-011-0490-1

Christensen, A. J., & Smith, T. W. (1995). Personality and patient adherence: Correlates of the five-factor model in renal dialysis. *Journal of Behavioral Medicine*, *18*, 305–313. https://doi.org/10.1007/BF01857875

Christensen, A. P., Golino, H., & Silvia, P. J. (2020). A psychometric network perspective on the validity and validation of personality trait questionnaires. *European Journal of Personality*. Advance online publication. https://doi.org/10.1002/per.2265

Corker, K. S., Oswald, F. L., & Donnellan, M. B. (2012). Conscientiousness in the classroom: A process explanation. *Journal of Personality*, *80*, 995–1028. https://doi.org/10.1111/j.1467-6494.2011.00750.x

Costantini, G., Richetin, J., Preti, E., Casini, E., Epskamp, S., & Perugini, M. (2019). Stability and variability of personality networks. A tutorial on recent developments in network psychometrics. *Personality and Individual Differences*, *136*, 68–78. https://doi.org/10.1016/j.paid.2017.06.011

Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco, CA: Jossey-Bass. https://doi.org/10.1177/109821408300400210

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. New York, NY: John Wiley & Sons. https://doi.org/10.1126/science.178.4067.1275

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302. https://doi.org/10.1037/h0040957

Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology*, *16*, 137–163. https://doi.org/10.1111/j.2044-8317.1963.tb00206.x

Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method. In *Flow and the foundations of positive psychology* (pp. 35–54). Dordrecht, Netherlands: Springer.

De Mauro, A., Greco, M., & Grimaldi, M. (2015). What is big data? A consensual definition and a review of key research topics. *AIP Conference Proceedings*, *1644*, 97–104. https://doi.org/10.1063/1.4907823

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, *55*, 78–87. https://doi.org/10.1145/2347736.2347755

Donnellan, M. B., & Robins, R. W. (2010). Resilient, overcontrolled, and undercontrolled personality types: Issues and controversies. *Social and Personality Psychology Compass*, *4*, 1070–1083. https://doi.org/10.1111/j.1751-9004.2010.00313.x

Epskamp, S., Rhemtulla, M., & Borsboom, D. (2017). Generalized network psychometrics: Combining network and latent variable models. *Psychometrika; Williamsburg*, *82*, 904–927. https://doi.org.ezproxy.rice.edu/10.1007/s11336-017-9557-x

Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality*, *51*, 360–392. https://doi.org/10.1111/j.1467-6494.1983.tb00338.x

Equal Employment Opportunity Commission (1978). Uniform guidelines on employee selection procedures. *Federal Register*, *43*, 38295–38309.

Erdmier, C., Hatcher, J., & Lee, M. (2016). Wearable device implications in the healthcare industry. *Journal of Medical Engineering & Technology*, *40*, 141–148. https://doi.org/10.3109/03091902.2016.1153738

Fan, J., & Li, R. (2006). Statistical challenges with high dimensionality: Feature selection in knowledge discovery. *ArXiv:Math*. Retrieved from. https://arxiv.org/abs/math/0602133

Fiedler, K. (2018). The creative cycle and the growth of psychological science. *Perspectives on Psychological Science*, *13*, 433–438. https://doi.org/10.1177/1745691617745651

Flake, J. K., & Fried, E. I. (2019). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *PsyArXiv. Advance online publication.*. https://doi.org/10.31234/osf.io/hs7wm

Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science*, *13*, 83–87. https://doi.org/10.1111/j.0963-7214.2004.00280.x

Foster, K., Schuh, S., & Zhang, H. (2013). The 2010 survey of consumer payment choice. *Research Reviews*, *20*, 113–118. https://doi.org/10.2139/ssrn.2564172

Friedman, J., Hastie, T., & Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, *9*, 432–441. https://doi.org/10.1093/biostatistics/kxm045

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, *35*, 137–144. https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gerlach, M., Farb, B., Revelle, W., & Amaral, L. A. N. (2018). A robust data-driven approach identifies four personality types across four large data sets. *Nature Human Behaviour*, *2*, 735–742. https://doi.org/10.1038/s41562-018-0419-z

Gladstone, J. J., Matz, S. C., & Lemaire, A. (2019). Can psychological traits be inferred from spending? Evidence from transaction data. *Psychological Science*, *30*, 1087–1096. https://doi.org/10.1177/0956797619849435

Golbeck, J., Robles, C., Edmondson, M., & Turner, K. (2011). *Predicting personality from Twitter*. 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, 149–156. https://doi.org/10.1109/PASSAT/SocialCom.2011.33

Golbeck, J., Robles, C., & Turner, K. (2011). Predicting personality with social media. *CHI EA'11: CHI'11 Extended Abstracts on Human Factors in Computing Systems*, 253–262. https://doi.org/10.1145/1979742.1979614

Hamaker, E. L., Nesselroade, J. R., & Molenaar, P. C. (2007). The integrated trait-state model. *Journal of Research in Personality*, *41*, 295–315. https://doi.org/10.1016/j.jrp.2006.04.003

Haqiqatkhah, M. M., & Tuerlinckx, F. (2019). Are we on the same page? Latent variable modeling suggests different nomothetic and idiographic factor structures for momentary affect. *PsyArXiv*. Advance online publication. https://doi.org/10.31234/osf.io/6wsgd

Harari, G. M., Lane, N. D., Wang, R., Crosier, B. S., Campbell, A. T., & Gosling, S. D. (2016). Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. *Perspectives on Psychological Science*, *11*, 838–854. https://doi.org/10.1177/1745691616650285

Hawkins, D. M. (2004). The problem of overfitting. *Journal of Chemical Information and Computer Sciences*, *44*, 1–12. https://doi.org/10.1021/ci0342472

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83. https://doi.org/10.1017/S0140525X0999152X

Huffman, S. (2018). The digital divide revisited: What is next? *Education*, *138*, 239–246. Retrieved from. https://search-ebscohost-com.ezproxy.rice.edu

Iacobelli, F., Gill, A. J., Nowson, S., & Oberlander, J. (2011). Large scale personality classification of bloggers. In S. D'Mello, A. Graesser, B. Schuller, & J.-C. Martin (Eds.), *Affective computing and intelligent interaction* (pp. 568–577). https://doi.org/10.1007/978-3-642-24571-8_71

Ihsan, Z., & Furnham, A. (2018). The new technologies in personality assessment: A review. *Consulting Psychology Journal: Practice and Research*, *70*, 147–166. https://doi.org/10.1037/cpb0000106

Ilies, R., & Judge, T. A. (2002). Understanding the dynamic relationships among personality, mood, and job satisfaction: A field experience sampling study. *Organizational Behavior and Human Decision Processes*, *89*, 1119–1139. https://doi.org/10.1016/S0749-5978(02)00018-3

Inoubli, W., Aridhi, S., Mezni, H., Maddouri, M., & Nguifo, E. M. (2018). An experimental survey on big data frameworks. *Future Generation Computer Systems*, *86*, 546–564. https://doi.org/10.1016/j.future.2018.04.032

Ippel, L., Kaptein, M. C., & Vermunt, J. K. (2019). Estimating multilevel models on data streams. *Psychometrika*, *84*, 41–64. https://doi.org/10.1007/s11336-018-09656-z

Jackson, L. A., Zhao, Y., Kolenic, A. III, Fitzgerald, H. E., Harold, R., & Von Eye, A. (2008). Race, gender, and information technology use: The new digital divide. *Cyberpsychology & Behavior*, *11*, 437–442. https://doi.org/10.1089/cpb.2007.0157

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning*. New York, NY: Springer. https://doi.org/10.1007/978-1-4614-7138-7

Jeong, I. C., Bychkov, D., & Searson, P. C. (2019). Wearable devices for precision medicine and health state monitoring. *IEEE Transactions on Biomedical Engineering*, *66*, 1242–1258. https://doi.org/10.1109/TBME.2018.2871638

Judge, T. A., Simon, L. S., Hurst, C., & Kelley, K. (2014). What I experienced yesterday is who I am today: Relationship of work motivations and behaviors to within-individual variation in the five-factor model of personality. *Journal of Applied Psychology*, *99*, 199–221. https://doi.org/10.1037/a0034485

Kelley, T. L. (1927). *Interpretation of educational measurements*. New York, NY: World Book Company.

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. https://doi.org/10.1037/a0039210

Kosinski, M., Wang, Y., Lakkaraju, H., & Leskovec, J. (2016). Mining big data to extract patterns and predict real-life outcomes. *Psychological Methods*, *21*, 493–506. https://doi.org/10.1037/met0000105

Kuhn, M. (2019). *caret: Classification and regression training*. R package version 6.0–84. https://CRAN.R-project.org/package=caret

Landers, R. N., Brusso, R. C., Cavanaugh, K. J., & Collmus, A. B. (2016). A primer on theory-driven web scraping: Automatic extraction of big data from the internet for use in psychological research. *Psychological Methods*, *21*, 475–492. https://doi.org/10.1037/met0000081

Laney, D. (2001, February 6). *3-D Data management: Controlling data volume, velocity, and variety*. Application Delivery Strategies by META Group Inc. Retrieved from https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf

Lenhart, A., Duggan, M., Perrin, A., Steppler, R., Rainie, L., & Parker, K. (2015). *Teens, social media, & technology overview 2015: Smartphones facilitate shifts in communication landscape for teens* (p. 48). Retrieved from https://www.pewresearch.org/wp-content/uploads/sites/9/2015/04/PI_TeensandTech_Update2015_0409151.pdf

Lindsay, C. (2005). Employability, services for unemployed job seekers and the digital divide. *Urban Studies*, *42*, 325–339. https://doi.org/10.1080/0042098042000316173

Lodge, J. M., Alhadad, S. S. J., Lewis, M. J., & Gašević, D. (2017). Inferring learning from big data: The importance of a transdisciplinary and multidimensional approach. *Technology, Knowledge and Learning*, *22*, 385–400. https://doi.org/10.1007/s10758-017-9330-3

Lüdtke, O., Roberts, B. W., Trautwein, U., & Nagy, G. (2011). A random walk down university avenue: Life paths, life events, and personality trait change at the transition to university life. *Journal of Personality and Social Psychology*, *101*, 620–637. https://doi.org/10.1037/a0023743

Mackert, M., Mabry-Flynn, A., Champlin, S., Donovan, E. E., & Pounders, K. (2016). Health literacy and health information technology adoption: The potential for a new digital divide. *Journal of Medical Internet Research*, *18*, 211–226. https://doi.org/10.2196/jmir.6349

Mansour, R. F. (2016). Understanding how big data leads to social networking vulnerability. *Computers in Human Behavior*, *57*, 348–351. https://doi.org/10.1016/j.chb.2015.12.055

Mardonova, M., & Choi, Y. (2018). Review of wearable device technology and its applications to the mining industry. *Energies*, *11*, 1–14. https://doi.org/10.3390/en11030547

Marler, W. (2018). Mobile phones and inequality: Findings, trends, and future directions. *New Media & Society*, *20*, 3498–3520. https://doi.org/10.1177/1461444818765154

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, *22*, 471–491. https://doi.org/10.1037/a0019227

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*, 175–215. https://doi.org/10.1111/j.1467-6494.1992.tb00970.x

Meade, A. W., Michels, L. C., & Lautenschlager, G. J. (2007). Are internet and paper-and-pencil personality tests truly comparable?: An experimental design measurement invariance study. *Organizational Research Methods*, *10*, 322–345. https://doi.org/10.1177/1094428106289393

Meers, K., Dejonckheere, E., Kalokerinos, E. K., Rummens, K., & Kuppens, P. (2019, June 12). *mobileQ: A free user-friendly application for collecting experience sampling data*. https://doi.org/10.31234/osf.io/ynj7e

Millsap, R. E. (2007). Invariance in measurement and prediction revisited. *Psychometrika*, *72*, 461–473. https://doi.org/10.1007/s11336-007-9039-7

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY: Taylor & Francis.

Minbashian, A., Wood, R. E., & Beckmann, N. (2010). Task-contingent conscientiousness as a unit of personality at work. *Journal of*

Applied Psychology, *95*, 793–806. https://doi.org/10.1037/a0020016

Molenaar, P. C., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science*, *18*, 112–117. https://doi.org/10.1111/j.1467-8721.2009.01619.x

Mønsted, B., Mollgaard, A., & Mathiesen, J. (2018). Phone-based metric as a predictor for basic personality traits. *Journal of Research in Personality*, *74*, 16–22. https://doi.org/10.1016/j.jrp.2017.12.004

Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *JAMA*, *309*, 1351–1352. https://doi.org/10.1001/jama.2013.393

Nosek, B. A., & Errington, T. M. (2020). What is replication? *PLoS Biology*, *18*, 1–8. https://doi.org/10.1371/journal.pbio.3000691

Ones, D. S., & Wiernik, B. M. (2018, October 10). *On "new" personality types*. Retrieved from https://www.siop.org/Research-Publications/Items-of-Interest/ArtMID/19366/ArticleID/1698/On-%E2%80%9CNew%E2%80%9D-Personality-Types

Perrin, A., & Anderson, M. (2019, April 10). *Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018*. Retrieved May 27, 2019, from Pew Research Center website: https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/

Quercia, D., Kosinski, M., Stillwell, D., & Crowcroft, J. (2011). *Our Twitter profiles, our selves: Predicting personality with Twitter*. 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, 180–185. https://doi.org/10.1109/PASSAT/SocialCom.2011.26

Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, *84*, 1236–1254. https://doi.org/10.1037/0022-3514.84.6.1236

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016a). Model-agnostic interpretability of machine learning. In B. Kim, D. M. Malioutov, & K. R. Varshney (Eds.), *Proceedings of the 2016 ICML Workshop on Human Interpretability in Machine Learning* (pp. 91–95). https://doi.org/10.1145/2939672.2939778

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016b). *"Why should I trust you?": Explaining the predictions of any classifier*. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144. https://doi.org/10.1145/2939672.2939778

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, *2*, 313–345. https://doi.org/10.1111/j.1745-6916.2007.00047.x

Rocher, L., Hendrickx, J. M., & de Montjoye, Y.-A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications*, *10*, 1–9. https://doi.org/10.1038/s41467-019-10933-3

Roethlisberger, F. J., & Dickson, W. J. (1939). *Management and the worker*. Oxford, England: Harvard University Press.

Rowsell, J., Morrell, E., & Alvermann, D. E. (2017). Confronting the digital divide: Debunking brave new world discourses. *Reading Teacher*, *71*, 157–165. https://doi.org/10.1002/trtr.1603

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*, 206–215. Retrieved from. https://arxiv.org/abs/1811.10154v3

Salgado, J. F., & Moscoso, S. (2003). Internet-based personality testing: Equivalence of measures and assesses' perceptions and reactions. *International Journal of Selection and Assessment*, *11*, 194–205. https://doi.org/10.1111/1468-2389.00243

Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., Shah, A., … Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS ONE*, *8*, 1–16. https://doi.org/10.1371/journal.pone.0073791

Settanni, M., Azucar, D., & Marengo, D. (2018). Predicting individual characteristics from digital traces on social media: A meta-analysis. *CyberPsychology, Behavior & Social Networking*, *21*, 217–228. https://doi.org/10.1089/cyber.2017.0384

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science*, *12*, 1123–1128. https://doi.org/10.1177/1745691617708630

Society for Industrial and Organizational Psychology (2019). Principles for the validation and use of personnel selection procedures. *Industrial and Organizational Psychology*, *11*, 1–97. https://doi.org/10.1017/iop.2018.195

Soini, S., Mustajoki, P., Eriksson, J. G., & Lahti, J. (2018). Personality traits associated with weight maintenance among successful weight losers. *American Journal of Health Behavior*, *42*, 78–84. https://doi.org/10.5993/AJHB.42.6.8

Stillwell, D. J., & Kosinski, M. (2020, May 1). *myPersonality project website*. Retrieved from http://mypersonality.org/

Storey, V. C., & Song, I.-Y. (2017). Big data technologies and management: What conceptual modeling can do. *Data & Knowledge Engineering*, *108*, 50–67. https://doi.org/10.1016/j.datak.2017.01.001

Strickhouser, J., Zell, E., & Krizan, Z. (2017). Does personality predict health and well-being? A metasynthesis. *Health Psychology*, *36*, 797–810. https://doi.org/10.1037/hea0000475

Sumner, C., Byers, A., Boochever, R., & Park, G. J. (2012). *Predicting dark triad personality traits from Twitter usage and a linguistic analysis of tweets*. 2012 11th International Conference on Machine Learning and Applications, 2, 386–393. https://doi.org/10.1109/ICMLA.2012.218

Sumner, C., Byers, A., & Shearing, M. (2011). Determining personality traits & privacy concerns from Facebook activity. *Black Hat Briefings*, *11*, 197–221.

Sutin, A. R., Ferrucci, L., Zonderman, A. B., & Terracciano, A. (2011). Personality and obesity across the adult life span. *Journal of Personality and Social Psychology*, *101*, 579–592. https://doi.org/10.1037/a0024286

Tay, L., Woo, S. E., Hickman, L., & Saef, R. (2020). Psychometric and validity issues in machine learning approaches to personality assessment: A focus on social media text mining. *European Journal of Personality*, *34*, 826–844. https://doi.org/10.1002/per.2290

Thorndike, E. L. (1904). *An introduction to the theory of mental and social measurements*. New York, NY: Columbia University Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*. Series B Methodological, *58*, 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Tskhay, K. O., & Rule, N. O. (2014). Perceptions of personality in text-based media and OSN: A meta-analysis. *Journal of Research in Personality*, *49*, 25–30. https://doi.org/10.1016/j.jrp.2013.12.004

Umaki, T. M., Umaki, M. R., & Cobb, C. M. (2012). The psychology of patient compliance: A focused review of the literature. *Journal of Periodontology*, *83*, 395–400. https://doi.org/10.1902/jop.2011.110344

Van Deursen, A. J., & Van Dijk, J. A. (2019). The first-level digital divide shifts from inequalities in physical access to inequalities in material access. *New Media & Society*, *21*, 354–375. https://doi.org/10.1177/1461444818797082

Van Volkom, M., Stapley, J. C., & Amaturo, V. (2014). Revisiting the digital divide: Generational differences in technology use in everyday life. *North American Journal of Psychology*, *16*, 557–574.

Wald, R., Khoshgoftaar, T., & Sumner, C. (2012). *Machine prediction of personality from Facebook profiles*. 2012 IEEE 13th

International Conference on Information Reuse Integration, 109–115. https://doi.org/10.1109/IRI.2012.6302998

Woo, S. E., Tay, L., Jebb, A. T., Ford, M. T., & Kern, M. L. (2020). Big data for enhancing measurement quality. In S. E. Woo, L. Tay, & R. W. Proctor (Eds.), *Big Data in Psychological Research* (pp. 59–85). American Psychological Association. https://doi.org/10.1037/0000193-004

Wright, A. G. C. (2014). Current directions in personality science and the potential for advances through computing. *IEEE Transactions on Affective Computing*, *5*, 292–296. https://doi.org/10.1109/TAFFC.2014.2332331

Yan, Y., Nie, J., Huang, L., Li, Z., Cao, Q., & Wei, Z. (2015). Is your first impression reliable? Trustworthy analysis using facial traits in portraits. In X. He, S. Luo, D. Tao, C. Xu, J. Yang, & M. A. Hasan (Eds.), *Multimedia modeling* (pp. 148–158). https://doi.org/10.1007/978-3-319-14442-9_13

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*, 1100–1122. https://doi.org/10.1177/1745691617693393

Yost, A. B., Behrend, T. S., Howardson, G., Darrow, J. B., & Jensen, J. M. (2019). Reactance to electronic surveillance: A test of antecedents and outcomes. *Journal of Business and Psychology*, *34*, 71–86. https://doi.org/10.1007/s10869-018-9532-2

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*, *112*, 1036–1040. https://doi.org/10.1073/pnas.1418680112

Youyou, W., Stillwell, D., Schwartz, H. A., & Kosinski, M. (2017). Birds of a feather do flock together: Behavior-based personality-assessment method reveals personality similarity among couples and friends. *Psychological Science*, *28*, 276–284. https://doi.org/10.1177/0956797616678187

Zimmermann, J., Woods, W. C., Ritter, S., Happel, M., Masuhr, O., Jaeger, U., … Wright, A. G. C. (2019). Integrating structure and dynamics in personality assessment: First steps toward the development and validation of a personality dynamics diary. *Psychological Assessment*, *31*, 516–531. https://doi.org/10.1037/pas0000625

DOI: 10.1002/per