The Personality Panorama: Conceptualizing Personality Through Big Behavioral Data

SCHOLARONE™
Manuscripts

Abstract

Personality Psychology has long been grounded in data typologies, particularly in the delineation of Behavioral, Life outcome, Informant-report, and Self-report (BLIS) sources of data from one another. Such data typologies are becoming obsolete in the face of new methods, technologies, and data philosophies. In this article, we discuss Personality Psychology's historical thinking about data, modern data theory's place in a personality psychology, and several qualities of big data that urge a rethinking of personality itself. We call for a move away from self-report questionnaires and a reprioritization of the study of behavior within personality science. With big data and behavioral assessment, we have the potential to witness the confluence of situated, seamlessly interacting psychological processes, forming an inclusive, dynamic, multi-angle view of personality. However, big behavioral data comes hand-in-hand with important ethical considerations, and our emerging ability to create a "personality panopticon" requires careful and thoughtful navigation. For our research to improve and thrive in partnership with new technologies, we must not only wield our new tools thoughtfully, but humanely. Through discourse and collaboration with other disciplines and the general public, we can foster mutual growth and ensure that humanity's burgeoning technological capabilities serve, rather than control, the public interest.

Keywords: big data; personality; psychometrics; behavior; ethics

The Personality Panorama:

Conceptualizing Personality Through Big Behavioral Data

Two of the most crucial features of scientific psychology in the 20th century were its grounding in objective behavior (as opposed to the introspectionism that preceded it) and its widespread embrace of statistics (including the law of large numbers). Nowhere was this more apparent than in the field of assessment, where empirical analyses of "large" samples drove the construction of tests of intelligence (Binet & Simon, 1916), vocational fit (Strong, 1934, 1935), psychopathology (Schiele et al., 1943), and personality (Gough, 1956).

Now, in the 21st century, we have entered the early stages of a new quantum shift as personality scientists grapple with the availability of Big Data and revisit paths that were forged by the "dustbowl empiricists" of yesteryear. In this article, we guide the reader away from entrenched deficiencies in the field's approach to personality data, identify the underpinnings shared by theories of both personality and data, and illustrate the ethical horizons that come into view as Psychology and Big Data grow evermore intertwined. In doing so, our goal is to help demystify big behavioral data, opening the gates to those who have otherwise come to understand "big data" as a towering colossus that is beyond the grasp of the typical psychologist.

Approaching a Modern Understanding of Personality Data

Personality psychology has long been a multi-method enterprise but, historically, the primary methods of personality assessment have been self- and observer ratings, supplemented by demographic characteristics and ostensibly objective measures of performance (Block, 1977). For Funder (2015), these performance measures are called "Behavioral" data which, together with Life, Informant, and Self-report, constitute the "BLIS-ful" territory of personality with

which the personality psychologists concerns themselves. Traditional typologies of personality data such as BLIS are intuitively appealing – they have a certain heuristic quality about them that have served as invaluable guideposts for the advancement of the field.

Nevertheless, typologies such as BLIS are limited and, particularly in the context of our modern technologies, leave us wanting. Separation between the "types" of personality data have always been blurry – for instance, self-reports of church attendance or employment history are commonly treated as proxy measures of one's behaviors and life outcomes. Moreover, the widely-issued personality data typologies are themselves incomplete, providing cramped quarters that often struggle to accommodate the rich, nuanced innovations in research methods and theory. An ever-growing list of data types could include, for example, "Things left behind" (Gosling et al., 2002), a person's "Words" as primary sources of data (Groom & Pennebaker, 2002), "Images" of faces (Todorov et al., 2005), and "Networks" of actions and persons (Staiano et al., 2012); these could jointly constitute a "TWIN" typology for contemporary Personality Psychology in a technologically enmeshed world.

Delightful acronyms aside, exhaustively labelling data types would be, at best, unproductive. Our growing ability to quantify directly observable human behavior (as well as behavioral traces) comes with a need to wholly rethink and reprioritize our understanding of personality data and the methods by which we collect it. Innovations in data analytics across academic disciplines have created opportunities that were unimaginable a mere decade ago. As with all scientific pursuits, new opportunities require us to question and update our traditional ideas about data and measurement within the field.

Rethinking the Role of Self-Report

For the personality psychologist, self-report questionnaires have historically been the most essential and widely used assessment tool. The ubiquitous role of questionnaires to the exclusion of other methods has come with considerable costs, particularly in areas where self-reports are unrelated to objective behaviors. Areas such as self-regulation (Cyders & Coskunpinar, 2011; B. Reynolds et al., 2006; Sharma et al., 2014), emotions (Mauss & Robinson, 2009; Wojcik et al., 2015), physical activity (Prince et al., 2008), sexuality (Stanton et al., 2015), and technology use (Boase & Ling, 2013; D. A. Ellis et al., 2019) are just a few domains where self-reports reveal a tenuous connection to objective measures of the same constructs. While self-reports are often taken as a face-valid representation of an empirical truth (e.g., a person who endorses "I am talkative" talks more than most) or an accurate self-perception ("I see myself as talkative"), they can also be construed simply as a signal with or without empirical significance (Meehl, 1945).

The challenges facing self-report data are particularly well-established and undermine their empirical value for personality psychologists (Furr, 2009a; Paulhus & Vazire, 2007; Pryor et al., 1977). Nevertheless, self-report questionnaires remain deeply entrenched in Social and Personality Psychology. Self-reports are often elevated to the status of "ground truth" measures whose validity is simply assumed, serving as the warped yardstick against which objective measures are compared (e.g., Sun, Schwartz, Son, Kern, & Vazire, 2019). While self-report questionnaires can be effective for assessing *subjective self-perceptions*[1], their widespread,

---

[1] We note that additional methods can be highly useful for discovering nuances that may not be reflected in existing and well-validated self-report questionnaires (e.g., Chung & Pennebaker, 2008; Stanton, Boyd, Pulverman, & Meston, 2015).

uncritical use valid measure of other domains is, at best, suboptimal (e.g., Healey, 2011); at its worst, the assumed "default" validity of self-report measures is deeply misleading and problematic (see, e.g., Cooper, 2019; D. A. Ellis et al., 2019).

In the 20th Century, self-report data were dominant, in part, because they were easier to collect than more objective behavioral measures. Today, however, the relative ease with which behavioral data can be collected has dispelled this advantage. The very nature of psychological data has changed as a function of the digital age, allowing us to reprioritize the role of behavioral assessment in personality science. Evolutions in personal computing and the creative use of online technology have pulled us into the age of a sensing, ubiquitous "internet of things", giving rise to a progressive digitalization of data that captures behavior in virtually every form. Today, raw behavioral data already exist *en masse*, and current technology has made it easier than ever to harness human activity in rich detail. We already have, and will continue to have, far more behavioral data than we know what to do with.

Beyond Typologies: Behavior as the Gold Standard

It been more than a decade since Baumeister, Vohs, and Funder (2007) and Furr (2009a, 2009b) lamented the lack of behavior in psychological research. Personality Science has been slow to reform, often citing the many costs and challenges associated with crafting and collecting objective assessments of naturalistic behavior (e.g., Hofmans, De Clercq, Kuppens, Verbeke, & Widiger, 2019; Wrzus & Mehl, 2015). Such concerns are reasonable and justified – the careful curation of behavioral data "from scratch" is no small feat. However, justification for the field's aversion to *existing* infrastructures and behavioral data are generally elusive; existing and easy-to-deploy technologies have caused issues of cost and data availability to largely evaporate. Considering the sheer abundance of already-digitized behavioral data, paired with the empirical

merits of objective measurement (Reiss & Sprenger, 2017), we urge psychologists to seriously

reconsider behavior – in all of its forms – as the gold standard for personality assessment.

A wide scope of behavior is of interest to the personality psychologist, ranging from the

atomic (e.g., a conditioned eyeblink) to symbolic, milestone acts comprised of important

supporting behaviors (e.g., getting married). The shifting landscape of human experience has

made the subtleties of behavioral *traces* particularly important: the physical and digital footprints

that human behavior leaves behind are a veritable goldmine of personality data (Lambiotte &

Kosinski, 2014). The images that a person chooses to share with others (Burdick et al., 2020;

Settanni et al., 2018), a person's words and other features of their verbal behavior (Boyd et al.,

2015; Golbeck, 2016; Hoover et al., 2018; Kern et al., 2014; Mitra et al., 2016; Park et al., 2015),

URL clicks (Lien et al., 2019; Tellakat et al., 2019), social behaviors (Adali & Golbeck, 2012;

Hilbig et al., 2015), and self-presentation behaviors (Liu et al., 2016; Segalin et al., 2017;

Shiramizu et al., 2019; Todorov et al., 2005; Walker et al., 2018) are no longer simply reflections

of personality – they are the critical raw material for understanding personality itself.

Personality psychology and the "big data" dimensions. The available bounty of

behavioral data is necessary, but not sufficient, for stepping into a new age of measurement.

Working with behavioral data requires us to think deeply about what information is captured in

behavior, and how it can be best understood as a reflection of the underlying psychology that

drives it (Rauthmann, 2017). If we are to reprioritize behavior, we must first remove the blinders

imposed by traditional data typologies (Goldstone & Lupyan, 2016). An updated, contemporary

understanding of big data theory is often treated as the dominion of the computational elite, yet

will be surprisingly familiar to those with training in psychometrics.[2]

Much like personality constructs, data itself differs along several continua or dimensions.

In the Big Data world, data is often characterized in dimensional terms along the "5 V's":

volume, variety, velocity, veracity and value (see, e.g., Yin & Kaynak, 2015). While the 5 V's of

big data are often discussed as a way to highlight computational challenges, these dimensions

can be mapped onto psychometric theory without much ado. By reframing modern personality

data in these terms, we hope to bridge a critical gap in Psychology's understanding of big data,

facilitating the rapid and organic uptake of Big Or Naturally Occurring Data Sets (BONDS;

Paxton & Griffiths, 2017) by personality psychologists.

Volume. Perhaps the most obvious feature of big behavioral data is that it exists in

extremely large quantities. Psychologists often think of data's volume in terms of *sample size* – a

critical aspect for generalizability in psychological research. Typical sample sizes in personality

research range from dozens to thousands; the number of participants in modern big data often

exceed those figures by several orders of magnitude. While personality psychologists may be

less plagued by issues of sample size than other subdisciplines (Soto, 2019), a consistent

underestimation of sample size adequacy persists in the field (Bakker et al., 2016). The sheer

volume of available behavioral data allows personality researchers expand their samples beyond

the problematic study of educated, Western, and industrialized individuals (Henrich et al., 2010),

---

[2] Many, if not most, concepts and techniques central to machine learning (including decision trees, linear
and logistic regression modeling, cross-validation, bootstrapping, and the "curse of dimensionality") have
long been familiar to those with training in psychometrics.

but also helps to ameliorate recent concerns with replication in the field (IntHout et al., 2015; Maxwell et al., 2015).

To many psychologists, the "volume" aspect of big data may often also seem to be its most daunting or prohibitive feature. As datasets have become larger by orders of magnitude, there has been a quantum shift from a psychology grounded in statistics, in which the essential challenge was to make sound inferences from too little data, to a psychology grounded in data science, in which the challenge is and will be to make meaningful inferences from too much data. However, this need not be the case – there is no singularly correct way to make use of big data, and making the most of big behavioral data often requires only minimal adaptation to become accessible to the social sciences. Procedures such as random subsampling and model averaging (e.g., bootstrap aggregation, or "bagging"; Breiman, 1996) require no specialized knowledge of computer science and build directly on the training that most psychologists receive in any graduate statistics course. In more computationally demanding situations, we note that partnership with computer, information, and data scientists can result in extremely fruitful collaborative outcomes.

Finally, we note that "big" and "little" data are not inherently at odds with one another. To the contrary, the insights gained from big data may be sharpened and further refined though careful, deliberate study in smaller samples, just as findings from small datasets may be generalized and refined through the use of big data. Echoing the sentiments of Blok and Pedersen (2014) and Stanton, Meston, and Boyd (2017), the pairing of big and small scale datasets can greatly enhance the quality of knowledge that can be gleaned from both.

Velocity. The second dimension of big behavioral data is its *velocity*, originally used to describe one's ability to process incoming data at (or near) real-time speeds. Implicit to the

concept of data's velocity is the speed at which data are sampled in the first place. In psychological terms, we can draw an intuitive parallel between *velocity* and *sampling rate*, which reflects our interest in personality dynamics as they unfold over various timescales. Data may be large due several sampling periods over an incredibly long timespan, such data collected over multiple lifetimes. Slowly unfolding personality processes can be assessed using the compilation of large, historical datasets, providing the opportunity to examine the concordance and stability between behavioral indicators of, for example, personality, politics, and culture (Iliev et al., 2016; Jordan et al., 2019; Obschonka et al., 2018).

Conversely, big data may also result from repeated, rapid sampling over the course of hours or minutes, as in the study of personality goals (Hudson & Fraley, 2015), states (Geukes et al., 2017; Power & Elliott, 2006; Wilt et al., 2011), and state transitions, among other complex dynamical processes (Vallacher et al., 2017).

As our ability to sample and aggregate data at varying levels of atomicity grows, big behavioral data will provide us with an increasingly "head to tail" view of personality. We seek to understand how personality operates from the momentary to lifetime to evolutionary scales, and everywhere in between. Our ability to acquire behavioral data with different velocities allows us to not only provide "snapshots" of personality at different levels of granularity, but begin to understand how a stream of personality "moments" flows into personality minutes, personality hours, days, months, years, and beyond.

Variety. *Variety* refers to the diversity of information contained within a dataset, both in the form of structured data (e.g., ordinal, nominal, integer, and ratio types) and unstructured data (e.g., raw text data, images, etc.). Variety of data can be expressed in several ways but, ultimately, gives us a sense of the range of phenomena that are (or can be) reflected in a given

dataset. For personality researchers, a multitude of data collection methods allow for the

automated, concurrent harvesting of data on tens of thousands of measures with minimal effort

(Landers, Brusso, Cavanaugh, & Collmus, 2016). Cell phones (Harari et al., 2017), wearable

technologies including cameras (Brown et al., 2017), microphones (Mehl, 2017), and

smartwatches (Hänsel et al., 2016), as well as geo-tagged location data (Mitchell et al., 2013) are

just a few of the additional data streams that have shown promise within the field, reflecting

various facets of behavior and psychology. Personality Psychology has long been a multimethod

enterprise; measures arising from behavior, introspection, and neuroscience all uniquely

contribute to our understanding of psychological phenomena (Anastasi & Urbina, 1997; Bosson

et al., 2000; Mauss & Robinson, 2009).

The inclusion of diverse indicators, extracted from structured and unstructured data at a

more or less granular level, allow for fine-tuned hypothesis formulation and testing (Carpita et

al., 2019; Landers et al., 2016). Recent work has shown, for example, that studies which include

a wider variety of behavioral indicators can more accurately and comprehensively predict

personality traits than those with less varied data (Azucar et al., 2018). Despite the difficulties

that can arise in terms of operationalization and statistical analysis, having varied indicators of

the same construct (e.g. number and content of daily posts, number of posted pictures with other

people/friends, answers and interactions to others' posts as components of the *extraverted*

personality) may be helpful in testing internal consistency reliability, convergent validity, as well

as discriminant validity of constructs themselves (e.g., Hair, Hult, Ringle, & Sarstedt, 2016;

Landers et al., 2016).

Importantly, some data are more varietal than others – the number of ways in which

reaction time data can be construed into psychological measures is typically far less than that of,

for example, image data, which can be deconstructed into thousands of metrics (e.g., color, variation, and the presence/absence of an inexhaustible list of objects/features). Researchers are now busy laying the psychometric groundwork for understanding how a combined variety of behavioral measures might reflect core psychological processes, however, much work remains to be done (Harari et al., 2017). Additionally, as technological progress occurs, the degree of *variety* that different types of data possess will grow – natural language processing, computer vision, and related technologies continue to provide us with new ways of quantifying data that simply did not exist until very recent history.

Veracity. The *veracity* of one's data, big or small, is akin to the notion of one of the most commonly discussed principles of measurement in Personality Psychology: *validity* (Balsis et al., 2015; A. Ellis, 1946; Krueger, 2005; Schmit et al., 1995). The idea that our measures actually tap into the constructs under study – and do so in a way that minimizes measurement error – is central to science itself and is no less true in a big data world.

As personality psychologists already know, our ability to predict one measure from another is insufficient evidence for the veracity of one's data.[3] An exclusive focus on sample sizes and prediction power can lead to serious mistakes — mistakes to which even big data experts are susceptible (O'Neil, 2016). In 2008, Google launched their Flu Trends platform, a "nowcasting" service designed to model and predict the spread of influenza by analyzing the

---

[3] Contemporary research at the intersection of personality, big data, and machine learning tends to fixate on the ability of machine learning algorithms to account for variance in self-report scores, taking *predictive validity* alone as evidence for successful operationalization and parameterization. As discussed above, the mere prediction of self-report measures is a problematic for Personality Psychology as a discipline, often resulting in nonsensical or misguided conclusions about the relationship between personality and behavior (see Boyd & Pennebaker, 2017).

words contained in the world's Google searches. Initially, the results seemed compelling and intuitive – regions where people searched for flu-like symptoms such as "headache" and "fever" appeared to also manifest higher influenza prevalence rates as reported by the Centers for Disease Control and Prevention (CDC). Despite Google's ability to accurately model the relationship between search terms and *past* CDC data, it was discovered that their predictors (i.e., the words in people's search terms) were compromised by several confounds, resulting in wild overestimations of real-time flu prevalence. Researchers eventually discovered that the platform was, in actuality, being driven largely by search terms that reflected more general seasonal patterns (e.g., "icy roads", "holidays", "high school basketball") that were not true signals of the actual influenza virus (Lazer et al., 2014; Lazer & Kennedy, 2015; O'Neil & Schutt, 2013). Google Flu Trends was shuttered in 2015 and has become an illustrative, cautionary tale of "big data hubris".

It would be unfair to belittle the fate of Google Flu Trends as the result of simple ignorance or negligence. However, one can speculate as to whether they should have consulted an extra psychometrician or two. The evaluation and improvement of big behavioral data's veracity is a task for which personality psychologists are especially well-suited, even in big data world (Bleidorn & Hopwood, 2019). Analytic teams that come from heavily computational backgrounds can benefit from insights provided by psychologists – insights that are crucial for reaching valid conclusions about the nature of personality from behavioral data (Boyd & Pennebaker, 2017). Personality psychologists receive extensive training in the science of ferreting out measurement issues – foundational examples include lessons learned in the measurement of personality disorders (e.g., Rosenthal & Hooley, 2010), aging (e.g., Nye, Allemand, Gosling, Potter, & Roberts, 2016), and workplace attainment (e.g., Hough, 2001). As

we dive ever deeper into new technologies, our enthusiasm must not overwhelm the fundamental need to establish the basic validity of our data (see, e.g., Sariyska & Montag, 2019). Put another way: as we continue our collective journey into the world of big data, we cannot forget to check the sureness of our footing with every additional step.

Value. The last V — *value* — arises from the collaboration and prudent treatment of the previous V's. The "value" of big behavioral data is perhaps the most subjective dimension, reflecting the potential of a dataset for achieving a desired goal. In the business and marketing worlds, big data's value is often defined by the degree to which it can be leveraged or transformed into revenue (e.g., Fan, Lau, & Zhao, 2015). In the field of healthcare, value can have multiple definitions, ranging from data's ability to enhance decision-making during diagnostics to improving the identification and prevention of fraud (Roski et al., 2014).

For psychologists, the value of big behavioral data remains to be fully articulated; the goals that we seek to attain will likely come in many forms. In practical terms, big data's value for personality research exists inasmuch as it can represent the breathtaking range of human behavior – behaviors which can be used to overcome many well-described shortcomings in the field (R. F. Baumeister et al., 2007; Furr, 2009a). In more ambitious terms, big behavioral data opens doors to capturing yet unknown facets of personality and other serendipities that go beyond the traditional, theoretical ground truths that we take for granted, such as the factorial or polarized structure of personality. If our goal is to better understand the human condition, big behavioral data affords us a richly *panoramic* view of personality – one that provides a comprehensive view of the individual across time and contexts.

The Personality Panorama

The relationship between knowledge and scientific theory is symbiotic: greater knowledge gives us the raw supplies used to construct better theories, and better theories are critical scaffolding that abstract, simplify, and generalize said knowledge (Keren & Schul, 2009). Traditional approaches to collecting and analyzing data have given us invaluable knowledge of personality, albeit in the form of highly fragmented snapshots (Mayer, 2005). Using big data, however, we can view behavior through a more comprehensive, 360-degree lens that captures personality from previously unseen angles.

In today's digital world, we can quantify not just what a person had for dinner, but which restaurant they went to, how they got there, and with whom they traveled – all of which conspire to give us an increasingly panoramic view of personality in action. By leveraging big behavioral data to create this *personality panorama* – an inclusive, broad-spectrum assessment of personality – we begin to cut across and transcend specialized research areas. In doing so, we see great potential for reconnecting our inquiries with those elemental, timeless questions that rest at the very heart of the human condition (see Mayer, 2007; see also Mayer & Allen, 2013).

The Multifaceted Nature of Behavior

A single slice of behavior is a wellspring of psychological insight, reflecting the combined output of a constellation of personality processes. A single sample of natural language, for example, constitutes not only a quantifiable behavior in the traditional sense (e.g., whether a person says that they are shy), but can be construed into concurrent measures of cognitive styles (Boyd & Pennebaker, 2017; Pennebaker et al., 2014), affect (Li et al., 2017; Saxbe et al., 2013), attachment (Borelli et al., 2017; Dunlop et al., 2019), motives (Boyd et al., 2019; Fetterman et al., 2015), maturity (Lanning et al., 2018), and even one's environment (Kramer et al., 2014;

Stella et al., 2018). Other sources of behavioral data are similarly multifaceted, reflecting a confluence of psychological processes in ways yet to be discovered.

By expanding a single behavioral data point into measures of an entire universe of psychological processes, we begin to see the potential for understanding how personality processes coalesce into personality itself (Boyd et al., 2019).[4] Consider someone who is drinking an alcoholic beverage: a behavior that, on its own, may be informative of their underlying personality in only the broadest sense (e.g., Cloninger, Sigvardsson, & Bohman, 1988). However, as addiction and substance abuse researchers already know, virtually every aspect of drinking behavior is critical for understanding it from a psychological perspective – as we begin to unpack additional information about the behavior, the various facets of one's personality come into increasingly clear focus. Are they drinking a single beer or pounding shots of hard liquor (e.g., Pardo, Aguilar, Molinuevo, & Torrubia, 2007)? Are they drinking alone at home, at a bar, or at an office party (e.g., Kahler, Read, Wood, & Palfai, 2003)? What time of day is it (e.g., Wilson, 1990)? Contextual data surrounding the behavior itself helps us to better triangulate whether a person is sensation-seeking or depressed, gregarious or fearing social rejection, drinking casually or struggling with alcohol dependence.

While a single post to a social media website can be immediately expanded into an array of psychological measures with text analytic methods, we urge readers to approach measurement with a mindset that is more inclusive, going beyond superficial, single-method perspectives. In

---

[4] The idea that a person's behavior reveals other important aspects of their psychology is hardly new. Watson (1913) keenly noted that behavior is necessarily yoked to cognition and emotion, and the notion of overlapping psychological processes is implicit to contemporary descriptions of personality data as well (Rauthmann, 2017).

addition to quantifying the *content* of a social media post, we can investigate data about *when* the post was made, whether the post was part of a previous conversation, *who* replied to the post, *where* the person was when they made the post, and so on. Going deeper still, each bit of contextual information – who, what, where, why, when, and how – can be thought of as the top-level of its own complex, subterranean network of important questions (e.g., not just networks of *who* replied, but measures of *how* they replied, timelines of *when* they replied, and so on, all deeply layered on top of each other) The uniquely multifaceted nature of behavior provides us with data that falls into multiple domains, and to various degrees, helping us to paint ever more panoramic pictures of personality, as it operates in the real world, in all of its beautiful complexity and intricacies.

Behavior, Contextualized

As a field, we have a long-standing tradition of acknowledging the importance of person × environment interactions, yet we are regularly reminded of our collective failure to act on this knowledge (Endler, 1982; Endler & Parker, 1992; K. J. Reynolds et al., 2010).[5] Nowhere has this shortcoming been more overlooked than in traditional assessment. A key limitation of self-report data is that it is typically interpreted without an appreciation for context. The responses of an MTurker completing a survey at home to earn a wage, a student seeking experimental credit at her university, and that of an individual seeking self-knowledge or counseling are all scored in the same way and without regard for the environmental forces that shape a person's data. Indeed,

---

[5] We note some exceptions that include the personality-style study of situations, as well as research attempting to understand what, psychologically speaking, constitutes a situation (Rauthmann et al., 2014; Rauthmann & Sherman, 2018). Fewer still (e.g., Sherman, Nave, & Funder, 2010) have dug deep into classically defined person-by-situation interactions.

it can be extremely difficult to capture *situated* personality using traditional approaches. Currently, even sophisticated studies of psychology in online settings treat people as isolated actors – we analyze images, language, and social media "likes" without an appreciation for the contextual factors that guide these behaviors.

In the world of big data, however, we find the explicit recognition that personality does not exist in a vacuum. In the digital realm, people do not merely behave, they *interact* with dynamic environmental factors that can also be measured – other people (Davidson et al., 2019; Pan et al., 2012), games (Canossa et al., 2015), webpages (Shobeiri et al., 2013; Turkyilmaz et al., 2015), smartphone apps (Miller, 2012), and so on. Social context parameters such as network size, centrality (position or role), and local transitivity (connectedness among friends), are associated with personality (Fang et al., 2015; Gosling et al., 2011; Staiano et al., 2012), and traits such as cooperation are at least in part emergent properties of context, dependent upon community structure (Apicella et al., 2012).

Indeed, the purpose, function, and impact of human behavior can often be understood more precisely when couched in relation to one's environment (e.g., Danescu-Niculescu-Mizil, West, Jurafsky, Leskovec, & Potts, 2013; Goldberg, Manian, Monroe, Potts, & Srivastava, 2015). In many ways, the study of humans through such *in situ* behaviors is no different than how we have come to understand personality in other species (Bremner-Harrison et al., 2018; Briffa & Greenaway, 2011; Tkaczynski et al., 2018). By accounting for the context in which human behavior occurs, we expand our scope of understanding the past, present, and future of personality at every level (Krause et al., 2010).

Panorama versus Panopticon: Ethical Substance in Personality Science

Over 200 years ago, social theorist and philosopher Jeremy Bentham (1791) detailed his architectural plans for a building that would allow a single manager to maximally observe their subjects. Bentham's design consisted of a central "inspection house" that provided near-complete surveillance capabilities as a means of social control – this arrangement was named "the panopticon", derived from the Greek word for "all seeing". Originally designed as a physical prison, Bentham's vision has since evolved into a symbol for social control at the societal level, driven by an omnipresent system that creates "a permanent account of individuals' behaviour" (Foucault, 1977) – a world where all behaviors, down to one's slightest movements, are ceaselessly observed and regulated. While the prospect of a panoramic, all-encompassing view of personality is both thrilling and inspiring, we recognize that the *panoptic* potential of our "personality panorama" is, quite literally, the stuff of dystopian nightmares (e.g., Mozur, 2018; Orwell, 1949).

Just as the application of factor analytic and other multivariate approaches to personality saw greater adoption following the proliferation of powerful mainframe computers, new technologies are leading not just to new research practices, but the possibility of previously unattainable perspectives on Murray's (1938) classic levels of personality — human nature, human differences, and human individuals. The spotlight on big data revalues a Baconian, inductive, *data-driven* method of achieving knowledge, echoing, perhaps, the so-called "dustbowl empiricism" of the 20th Century (see, e.g., Anderson, 2008). This earlier empiricism facilitated a revolution in measurement practices yet, importantly, came with widespread concerns regarding privacy (see Bier, 1980); shadows of both can still be seen today (Matz et al.,

2017). Our developing ability to make comprehensive assessments of a person's psychology carries with it heavy questions over how this power will be used, and who will be using it.

The significance of personality privacy. Concerns about personality research have been raised for more than half a century (e.g., Anastasi, 1967). The poet Kenneth Rexroth once decried the indignity of personality assessment; a Senate subcommittee investigated the extent to which administration of the MMPI violated privacy (Brayfield, 1965; Rexroth, 1959). New data technologies have acted as a powerful accelerant to concerns about the ethics of widespread, constant assessment. A primary focus has been on privacy and involuntary measurement, as in surveillance and the use of digital traces without informed consent (Paxton & Griffiths, 2017; Tufekci, 2019). But privacy is a concern for voluntary measures as well – compromises of data security have rendered assurances of "confidentiality" effectively meaningless. Data breaches include the public release of raw, non-anonymized responses, as in the posting of data from 70,000 users of *OKCupid* to the Open Science Framework (OSF) platform (Zimmer, 2018).

Of still greater concern is that even supposedly "anonymized" data are vulnerable as well. In one well-known example, Netflix account-holders were identified by linking their movie preferences to additional data scraped from a public source (Narayanan & Shmatikov, 2008); another involves the now-infamous, intentional release of insufficiently de-identified user data by America Online (Barbaro & Zeller, 2006). Elsewhere, the secure anonymization of personality data is likely to be even more challenging for several reasons. First, inferences about the identity of individuals can be made with a high degree of confidence from just a few fragments of information (Sweeney, 2005). As the amount of information about each respondent increases, the likelihood that anonymity can be maintained will drop further still. Second, because trait scores can be predicted by patterns of online social behavior (Bachrach et al., 2012;

Youyou et al., 2015), and because there is value when different datasets are joined (intellectual

capital, new knowledge), measures of personality will be increasingly linked with other forms of

data (Oboler et al., 2012). Third, many large-scale digital datasets will have been collected by

third-parties with limited ethical oversight (Thompson & Warzel, 2019), and personality scholars

will frequently have little control over how the data have been collected or maintained. Finally,

because personality data are and will continue to be valuable to such varied stakeholders as

financial institutions, product vendors, healthcare providers, intelligence organizations, potential

landlords, and prospective romantic partners, there will be a thriving market for personality-

relevant data (McMullan, 2015; Wu, 2019).

Matters of "personality privacy" are critical not just because they involve values that are

deeply felt by many around us, but also because threats to privacy have a range of social and

ethical consequences (see Custers et al., 2017, for a comprehensive treatment). If personality

includes identity and reputation (Hogan et al., 1996; McAbee & Connelly, 2016), then an

increase in the sphere of reputation implies a reduction in the private and personal. If intimacy

connotes privacy and exclusion, then a threat to privacy is a threat to intimacy. Similarly, if an

increase in the sphere of the public self implies a reduction in that of the private self – if our

concern is only how we appear to others, and not the "authentic behavioural expression of self"

(Maltby et al., 2012) – then threats to privacy compromise human authenticity. Further, threats to

privacy may exacerbate social inequalities (Shokri et al., 2011), including that between genders.

Privacy is, in part, a feminist issue, with concerns about victimization felt particularly widely

among women (Chang, 2019). De-anonymized digital traces can leave individuals saddled with a

reputation determined by their worst moments rather than by their best ones; many traditional

self-report measures, such as the California Psychological Inventory (CPI; Gough & Bradley, 1996), are expressly concerned with capturing the *extremes* of a personal profile.

Autonomy, including measurement and experimentation without consent, is a broad, but related, ethical concern (Gleibs, 2014). Here, the most dramatic and arguably consequential violation of autonomy occurred when Cambridge Analytica scraped the online activity of individuals whose friends had participated in a study of personality, then leveraged this further to assess the personality of all or nearly all American adults. Cambridge Analytica, Facebook, and other data-rich firms such as Experian who engage in practices such as "shadow profiling" – the nonconsensual profiling of individuals through their social connections (Garcia, 2017; Quodling, 2018) – offer personality-based manipulation. At the individual level, the selective tailoring of messages attenuates, at the very least, freedom of choice. At a societal level, it can lead to manipulations in consumer and political behavior, including voting decisions (Bond et al., 2012; Confessore & Hakim, 2017; Corbyn, 2012; Matz et al., 2016) – a pattern that threatens both democracy and informed self-determination.

Surmounting ethical challenges through collaboration and discourse. The history of Psychology is steeped in discussions that turn a critical gaze toward our ethics, pushing us to consider what our ethics do – and *should* – look like (e.g., Brierley, 1920; Coxe, 1940; Gluck, 1997; Krasner, 1962; Marks, 2018; Meerloo, 1956; Powell, Digdon, Harris, & Smithson, 2014; Prilleltensky, 1989; Seligman, 2018; Zimbardo, 1973). In the widespread discussion on ethics and big data, there is, of course, the recognized potential for societal benefits in the new age of personality measurement. Many symptoms of depression and other psychopathologies are already observable on social media using current methods (Birnbaum et al., 2017; Guntuku et al., 2017; Hassanpour et al., 2019; Reece & Danforth, 2017; Stanton et al., 2017; Thorstad & Wolff,

2019); it would be a grave mistake to ignore the reduction in human suffering that could be afforded by faster and more accurate diagnosis and treatment. But a utilitarian framework such as this may be insufficient, not just because the distribution of winners and losers is dramatically uneven, but because it conflicts with a Kantian view in which one does not use others as means to an end (Celie & Paris, 2019; Tunick, 2014). Privacy is fraught, for secrecy can destroy as well as elevate human dignity (Martin, 2019). Test-takers, social network-ers, and app users must be protected, and it is a fundamental right for them to have control and decision-making power over their "digital" lives.[6]

The fine line between the personality panorama and the personality panopticon is defined less by *what* data we collect, and more by *how* we use and protect it. Big data is not magical, and conducting research in a big data world requires the same elements as all other research: people, tools, questions, and theory (Kitchin, 2014). While there are no easy answers to the current (and yet to come) ethical concerns that accompany the personality panorama, there are uncontroversial processes by which we can perpetually operate and self-reflect. Chief among these practices are *collaboration* and *discourse* with others: scholars from various disciplines and backgrounds, the general public, and policymakers all have unique perspectives that continually shape our approach to human research (Kindon et al., 2007). As psychologists increasingly adopt big data methods and tools, we must collaborate with other specialists in order to maximally reap

---

[6] In Europe, these and related concerns about privacy, ethics, and data ownership have given rise to the General Data Protection Regulation (GDPR; European Union, 2016), which, among other features, grants people some degree of informational self-determination, provides individuals with the right of data erasure, and unifies data regulation across the European Union. However, the effectiveness of such measures for protecting individuals (versus second- and third-party stakeholders) remains to be seen (Roberts, 2018).

the value of each contributor's domain expertise (Harlow & Oswald, 2016). Just as most psychologists are ill-equipped to deal with the computational hurdles of big data, most computer scientists receive little to no training in the empirical study of affect, behavior, and cognition. It is precisely the non-overlapping expertise and unique vantages offered by people with diverse backgrounds, experiences, and perspectives that foster compatibility and fruitful collaboration.

Critically, and despite much hand-wringing in the popular press, many of today's prominent ethical debates are hardly new, and there is much be gained by remaining open to the perspectives that non-psychologists bring to the table. From a legal perspective, issues of privacy have been discussed extensively for well over 100 years (Warren & Brandeis, 1890). Philosophical discourse on topics such as privacy and self-determination has a history so well-seasoned that it makes a mere century of legal scholarship seem quaint by comparison (see Negley, 1966). Importantly, big data methods and technologies have not been engineered in an ethical vacuum, released to world without concern for consequences. Joseph Redmon, creator of the popular YOLO object detection algorithm, recently explained his decision to cease research on computer vision methods due to the high potential for abuse (Yuan, 2020), and many conferences now ask authors to include statements about the societal consequences of their work (e.g., Lin et al., 2020). Nevertheless, research on data privacy has blossomed in partnership with big data methodologies; the study of topics such as deidentification, data masking, and disclosure risk are more active than ever (see Torra, 2017).

In the same way that psychologists are learning from and borrowing methods in other fields, we can learn and grow as a result of their ethical discourse as well. Computational and social sciences are both deeply embedded within rich ethical traditions – traditions that are mutually informative. Drew (2016), for example, provides a brilliant overview of the ethical

framework that grew from the Government Data Science Partnership, drawing from insights gained during deliberative public workshops and surveys. In her overview, Drew details principles such as data security, accountability, and articulation of public need/benefit – principles that are highly familiar to psychologists, but whose application to big data expands beyond traditional ethical research guidelines (Manzo & Brightbill, 2007; Sugiura et al., 2017).

Through collaboration and continued discourse between fields, more generalized ethical principles have already begun to emerge for big data researchers – principles that are broadly applicable to researchers and professionals across disciplines (Hand, 2018). Just as collaboration and discussion are necessary for the application of new methods, technologies, and empirical frameworks, the same processes will continue to result in evolving ethical frameworks that prioritize factors such as integrity, accountability, and fairness. Indeed, psychological research on precisely such topics leaves us nicely positioned to take charge of shaping the future ethical landscape (Cummings & Day, 2019).

Conclusions

The big data revolution opens up new ways of conceptualizing personality as a construct as well as personality measurement methods. The imaginable potential, from both an academic and social point of view, is endless. Particularly through the use of behavioral measures, which not only complement but supersede many traditional methods, we can better explore and understand personality in all its forms and manifestations. As our understanding of behavior—personality links continues to develop in the context of interdisciplinary collaboration and evolving methods, we can confidently move forward into a new age of measurement.

We can imagine a not-too-distant future where the average person has unprecedented access to psychological self-knowledge – self-knowledge attained through an individualized

personality panorama – that helps us choose how to live our lives, both individually and as a society. The personality panorama afforded to us through the use of big behavioral data comes with important ethical questions – questions that can be answered through open discourse and collaboration other scholars, policy makers, and the general public. Personality Psychology should not pay for new wisdom with its morality; it is our very *obligation* to be mindful of the line that separates the panoramic from the panoptic. It is time for Psychology to take a seat at the "big data" table and not only to benefit from, but also to help shape, the future study of the human condition.

References

Adali, S., & Golbeck, J. (2012). Predicting personality with social behavior. *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 302–309. https://doi.org/10.1109/ASONAM.2012.58

Anastasi, A. (1967). Psychology, psychologists, and psychological testing. *American Psychologist*, *22*(4), 297–306. https://doi.org/10.1037/h0024521

Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Prentice Hall.

Anderson, C. (2008, June 23). The end of theory: The data deluge makes the scientific method obsolete. *Wired*. https://www.wired.com/2008/06/pb-theory/

Apicella, C. L., Marlowe, F. W., Fowler, J. H., & Christakis, N. A. (2012). Social networks and cooperation in hunter-gatherers. *Nature*, *481*(7382), 497–501. https://doi.org/10/fz3v4v

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

Bakker, M., Hartgerink, C. H. J., Wicherts, J. M., & van der Maas, H. L. J. (2016). Researchers' intuitions about power in psychological research. *Psychological Science*, *27*(8), 1069–1077. https://doi.org/10.1177/0956797616647519

Balsis, S., Cooper, L. D., & Oltmanns, T. F. (2015). Are informant reports of personality more internally consistent than self reports of personality? *Assessment*, *22*(4), 399–404. https://doi.org/10.1177/1073191114556100

Barbaro, M., & Zeller, T. (2006, August 9). A face is exposed for AOL searcher no. 4417749. *The New York Times*. https://www.nytimes.com/2006/08/09/technology/09aol.html

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science*, *2*(4), 396–403. https://doi.org/10.1111/j.1745-6916.2007.00051.x

Bentham, J. (1791). *Panopticon; Or, the Inspection-House* (2008 Reprinting). Dodo Press.

Bier, W. C. (1980). *Privacy, a vanishing value?* Fordham Univ Press.

Binet, A., & Simon, Th. (1916). *The development of intelligence in children (The Binet-Simon Scale)*. Williams & Wilkins Co. https://doi.org/10.1037/11069-000

Birnbaum, L. M., Ernala, K. S., Rizvi, F. A., De Choudhury, M., & Kane, M. J. (2017). A collaborative approach to identifying social media markers of schizophrenia by employing machine learning and clinical appraisals. *J Med Internet Res*, *19*(8), e289. https://doi.org/10.2196/jmir.7956

Bleidorn, W., & Hopwood, C. J. (2019). Using machine learning to advance personality assessment and theory. *Personality and Social Psychology Review*, *23*(2), 190–203. https://doi.org/10.1177/1088868318772990

Blok, A., & Pedersen, M. A. (2014). Complementary social science? Quali-quantitative experiments in a Big Data world. *Big Data & Society*, *1*(2), 2053951714543908. https://doi.org/10.1177/2053951714543908

Boase, J., & Ling, R. (2013). Measuring mobile phone use: Self-report versus log data. *Journal of Computer-Mediated Communication*, *18*(4), 508–519. https://doi.org/10.1111/jcc4.12021

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H.

(2012). A 61-million-person experiment in social influence and political mobilization.

Nature, *489*(7415). https://doi.org/10.1038/nature11421

Borelli, J. L., Ramsook, K. A., Smiley, P., Bond, D. K., West, J. L., & Buttitta, K. H. (2017).

Language matching among mother-child dyads: Associations with child attachment and

emotion reactivity. *Social Development*, *26*(3), 610–629.

https://doi.org/10.1111/sode.12200

Bosson, J. K., Swann, W. B., & Pennebaker, J. W. (2000). Stalking the perfect measure of

implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality*

and Social Psychology, *79*(4), 631–643. https://doi.org/10.1037/0022-3514.79.4.631

Boyd, R. L., Pasca, P., & Conroy-Beam, D. (2019). You're only Jung once: Building generalized

motivational systems theories using contemporary research on language. *Psychological*

Inquiry, *30*(2), 93–98. https://doi.org/10.1080/1047840X.2019.1633122

Boyd, R. L., & Pennebaker, J. W. (2017). Language-based personality: A new approach to

personality in a digital world. *Current Opinion in Behavioral Sciences*, *18*, 63–68.

https://doi.org/10.1016/j.cobeha.2017.07.017

Boyd, R. L., Wilson, S. R., Pennebaker, J. W., Kosinski, M., Stillwell, D. J., & Mihalcea, R.

(2015). Values in words: Using language to evaluate and understand personal values.

Proceedings of the Ninth International AAAI Conference on Web and Social Media, 31–

40. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10482

Brayfield, A. H. (1965). Testimony before the Senate Subcommittee on Constitutional Rights of

the Committee on the Judiciary. June 7-10, 1965, Washington, D. C. *The American*

Psychologist, *20*(11), 888–954.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140.

https://doi.org/10.1023/A:1018054314350

Bremner-Harrison, S., Cypher, B. L., Van Horn Job, C., & Harrison, S. W. R. (2018). Assessing

personality in San Joaquin kit fox in situ: Efficacy of field-based experimental methods

and implications for conservation management. *Journal of Ethology*, *36*(1), 23–33.

https://doi.org/10.1007/s10164-017-0525-9

Brierley, S. S. (1920). The present attitude of employees to Industrial Psychology. *British*

Journal of Psychology, *10*(2), 210–227. https://doi.org/10.1111/j.2044-

8295.1920.tb00020.x

Briffa, M., & Greenaway, J. (2011). High in situ repeatability of behaviour indicates animal

personality in the beadlet anemone actinia equina (cnidaria). *PLOS ONE*, *6*(7), e21963.

https://doi.org/10.1371/journal.pone.0021963

Brown, N. A., Blake, A. B., & Sherman, R. A. (2017). A snapshot of the life as lived: Wearable

cameras in social and personality psychological science. *Social Psychological and*

Personality Science, *8*(5), 592–600. https://doi.org/10/gbwrtw

Burdick, L., Mihalcea, R., Boyd, R. L., & Pennebaker, J. W. (2020). Analyzing connections

between user attributes, images, and text. *Cognitive Computation*.

https://doi.org/10.1007/s12559-019-09695-3

Canossa, A., Badler, J. B., El-Nasr, M. S., Tignor, S., & Colvin, R. C. (2015). In your face(t):

Impact of personality and context on gameplay behavior. *Proceedings of the 10th*

International Conference on the Foundations of Digital Games, 1–10.

Carpita, M., Ciavolino, E., & Pasca, P. (2019). Exploring and modelling team performances of

the Kaggle European Soccer database. *Statistical Modelling*, *19*(1), 74–101.

https://doi.org/10.1177/1471082X18810971

Celie, K.-B., & Paris, J. J. (2019). Respect for personhood: Concrete implications of a

philosophical misunderstanding. *Clinical Ethics*, *14*(3), 146–150.

https://doi.org/10.1177/1477750919851057

Chang, E. (2019, April 10). What women know about the internet. *The New York Times*.

https://www.nytimes.com/2019/04/10/opinion/privacy-feminism.html

Chung, C. K., & Pennebaker, J. W. (2008). Revealing dimensions of thinking in open-ended self-

descriptions: An automated meaning extraction method for natural language. *Journal of

Research in Personality*, *42*(1), 96–132. https://doi.org/10.1016/j.jrp.2007.04.006

Cloninger, C. R., Sigvardsson, S., & Bohman, M. (1988). Childhood personality predicts alcohol

abuse in young adults. *Alcoholism: Clinical and Experimental Research*, *12*(4), 494–505.

https://doi.org/10.1111/j.1530-0277.1988.tb00232.x

Confessore, N., & Hakim, D. (2017, March 6). Data firm says 'secret sauce' aided Trump; many

scoff. *The New York Times*. https://www.nytimes.com/2017/03/06/us/politics/cambridge-

analytica.html

Cooper, C. (2019). Pitfalls of personality theory. *Personality and Individual Differences*, *151*,

109551. https://doi.org/10.1016/j.paid.2019.109551

Corbyn, Z. (2012). Facebook experiment boosts US voter turnout. *Nature News*.

https://doi.org/10.1038/nature.2012.11401

Coxe, W. W. (1940). Tentative report of the Committee on Professional Ethics, American

Association for Applied Psychology, September, 1939. *Journal of Consulting

Psychology*, *4*(1), 24–26. https://doi.org/10.1037/h0050914

Cummings, J. A., & Day, T. E. (2019). But what do participants want? Comment on the "Data

Sharing in Psychology" special section (2018). *American Psychologist*, *74*(2), 245–247.

https://doi.org/10.1037/amp0000408

Custers, B., La Fors, K., Jozwiak, M., Esther, K., Bachlechner, D., Friedewald, M., & Aguzzi, S.

(2017). *Lists of ethical, legal, societal and economic issues of big data technologies*

(SSRN Scholarly Paper ID 3091018). Social Science Research Network.

https://papers.ssrn.com/abstract=3091018

Cyders, M. A., & Coskunpinar, A. (2011). Measurement of constructs using self-report and

behavioral lab tasks: Is there overlap in nomothetic span and construct representation for

impulsivity? *Clinical Psychology Review*, *31*(6), 965–982.

https://doi.org/10.1016/j.cpr.2011.06.001

Danescu-Niculescu-Mizil, C., West, R., Jurafsky, D., Leskovec, J., & Potts, C. (2013). No

country for old members: User lifecycle and linguistic change in online communities.

Proceedings of the 22Nd International Conference on World Wide Web, 307–318.

https://doi.org/10.1145/2488388.2488416

Davidson, B. I., Jones, S. L., Joinson, A. N., & Hinds, J. (2019). The evolution of online

ideological communities. *PLOS ONE*, *14*(5), e0216932.

https://doi.org/10.1371/journal.pone.0216932

Drew, C. (2016). Data science ethics in government. *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences*, *374*(2083). https://doi.org/10.1098/rsta.2016.0119

Dunlop, W. L., Karan, A., Wilkinson, D., & Harake, N. (2019). Love in the first degree: Individual differences in first-person pronoun use and adult romantic attachment styles. *Social Psychological and Personality Science*, 1948550619847455. https://doi.org/10.1177/1948550619847455

Ellis, A. (1946). The validity of personality questionnaires. *Psychological Bulletin*, *43*(5), 385– 440. https://doi.org/10.1037/h0055483

Ellis, D. A., Davidson, B. I., Shaw, H., & Geyer, K. (2019). Do smartphone usage scales predict behavior? *International Journal of Human-Computer Studies*, *130*, 86–92. https://doi.org/10.1016/j.ijhcs.2019.05.004

Endler, N. S. (1982). Interactionism: A personality model, but not yet a theory. *Nebraska Symposium on Motivation*, 155–200.

Endler, N. S., & Parker, J. D. A. (1992). Interactionism revisited: Reflections on the continuing crisis in the personality area. *European Journal of Personality*, *6*(3), 177–198. https://doi.org/10.1002/per.2410060302

Fan, S., Lau, R. Y. K., & Zhao, J. L. (2015). Demystifying big data analytics for business intelligence through the lens of marketing mix. *Big Data Research*, *2*(1), 28–32. https://doi.org/10.1016/j.bdr.2015.02.006

Fang, R., Landis, B., Zhang, Z., Anderson, M. H., Shaw, J. D., & Kilduff, M. (2015). Integrating personality and social networks: A meta-analysis of personality, network position, and

work outcomes in organizations. *Organization Science*, *26*(4), 1243–1260.

https://doi.org/10.1287/orsc.2015.0972

Fetterman, A. K., Boyd, R. L., & Robinson, M. D. (2015). Power versus affiliation in political

ideology. *Personality and Social Psychology Bulletin*, *41*(9), 1195–1206.

https://doi.org/10.1177/0146167215591960

Foucault, M. (1977). *Discipline and punish: The birth of the prison*. Pantheon Books.

Funder, D. C. (2015). *The personality puzzle* (7th ed.). W. W. Norton & Company, Inc.

Furr, R. M. (2009a). Personality psychology as a truly behavioural science. *European Journal of*

Personality, *23*(5), 369–401. https://doi.org/10.1002/per.724

Furr, R. M. (2009b). The study of behaviour in personality psychology: Meaning, importance

and measurement. *European Journal of Personality*, *23*(5), 437–453.

https://doi.org/10.1002/per.726

Garcia, D. (2017). Leaking privacy and shadow profiles in online social networks. *Science*

Advances, *3*(8). https://doi.org/10.1126/sciadv.1701172

Geukes, K., Nestler, S., Hutteman, R., Küfner, A. C. P., & Back, M. D. (2017). Trait personality

and state variability: Predicting individual differences in within- and cross-context

fluctuations in affect, self-evaluations, and behavior in everyday life. *Journal of Research*

in Personality, *69*, 124–138. https://doi.org/10.1016/j.jrp.2016.06.003

Gleibs, I. H. (2014). Turning virtual public spaces into laboratories: Thoughts on conducting

online field studies using social network sites. *Analyses of Social Issues and Public*

Policy, *14*(1), 352–370. https://doi.org/10.1111/asap.12036

Gluck, J. P. (1997). Harry F. Harlow and animal research: Reflection on the ethical paradox.

Ethics & Behavior, *7*(2), 149–161. https://doi.org/10.1207/s15327019eb0702_6

Golbeck, J. A. (2016). Predicting personality with social media. *AIS Transactions on Replication Research*, *2*(2), 1–10.

Goldberg, A., Manian, G., Monroe, W., Potts, C., & Srivastava, S. B. (2015). Fitting in or standing out? The tradeoffs of structural and cultural embeddedness. *Academy of Management Proceedings*, *2015*(1), 12263. https://doi.org/10.5465/AMBPP.2015.12263abstract

Goldstone, R. L., & Lupyan, G. (2016). Discovering psychological principles by mining naturally occurring data sets. *Topics in Cognitive Science*, *8*(3), 548–568. https://doi.org/10.1111/tops.12212

Gosling, S. D., Augustine, A. A., Vazire, S., Holtzman, N., & Gaddis, S. (2011). Manifestations of personality in online social networks: Self-reported facebook-related behaviors and observable profile information. *Cyberpsychology, Behavior, and Social Networking*, *14*(9), 483–488. https://doi.org/10.1089/cyber.2010.0087

Gosling, S. D., Ko, S. J., Mannarelli, T., & Morris, M. E. (2002). A room with a cue: Personality judgments based on offices and bedrooms. *Journal of Personality and Social Psychology*, *82*(3), 379–398. https://doi.org/10.1037/0022-3514.82.3.379

Gough, H. G. (1956). *California Psychological Inventory*. Consulting Psychologists Press.

Gough, H. G., & Bradley, P. (1996). *CPI Manual*. Consulting Psychologists Press.

Groom, C. J., & Pennebaker, J. W. (2002). Words. *Journal of Research in Personality*, *36*, 615–621.

Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2017). Detecting depression and mental illness on social media: An integrative review. *Current Opinion in Behavioral Sciences*, *18*, 43–49. https://doi.org/10.1016/j.cobeha.2017.07.005

Hair, J., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2016). *A primer on partial least squares structural equation modeling* (Second edition). SAGE Publications, Inc.

Hand, D. J. (2018). Aspects of data ethics in a changing world: Where are we now? *Big Data*, *6*(3), 176–190. https://doi.org/10.1089/big.2018.0083

Hänsel, K., Alomainy, A., & Haddadi, H. (2016). Large scale mood and stress self-assessments on a smartwatch. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct - UbiComp '16*, 1180–1184. https://doi.org/10/ggcgcp

Harari, G. M., Müller, S. R., Aung, M. S., & Rentfrow, P. J. (2017). Smartphone sensing methods for studying behavior in everyday life. *Current Opinion in Behavioral Sciences*, *18*, 83–90. https://doi.org/10.1016/j.cobeha.2017.07.018

Harlow, L. L., & Oswald, F. L. (2016). Big data in psychology: Introduction to the special issue. *Psychological Methods*, *21*(4), 447–457. https://doi.org/10.1037/met0000120

Hassanpour, S., Tomita, N., DeLise, T., Crosier, B., & Marsch, L. A. (2019). Identifying substance use risk based on deep neural networks and Instagram social media data. *Neuropsychopharmacology*, *44*(3), 487–494. https://doi.org/10.1038/s41386-018-0247-x

Healey, J. (2011). Recording affect in the field: Towards methods and metrics for improving ground truth labels. In S. D'Mello, A. Graesser, B. Schuller, & J.-C. Martin (Eds.), *Affective Computing and Intelligent Interaction* (pp. 107–116). Springer. https://doi.org/10.1007/978-3-642-24600-5_14

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*(2–3), 61–83; discussion 83-135. https://doi.org/10.1017/S0140525X0999152X

Hilbig, B. E., Thielmann, I., Hepp, J., Klein, S. A., & Zettler, I. (2015). From personality to

altruistic behavior (and back): Evidence from a double-blind dictator game. *Journal of

Research in Personality*, *55*, 46–50. https://doi.org/10.1016/j.jrp.2014.12.004

Hofmans, J., De Clercq, B., Kuppens, P., Verbeke, L., & Widiger, T. A. (2019). Testing the

structure and process of personality using ambulatory assessment data: An overview of

within-person and person-specific techniques. *Psychological Assessment*, *31*(4), 432–

443. https://doi.org/10.1037/pas0000562

Hogan, R., Hogan, J., & Roberts, B. W. (1996). Personality measurement and employment

decisions: Questions and answers. *American Psychologist*, *51*(5), 469–477.

https://doi.org/10.1037/0003-066X.51.5.469

Hoover, J., Dehghani, M., Johnson, K., Iliev, R., & Graham, J. (2018). Into the wild: Big data

analytics in moral psychology. In *Atlas of moral psychology* (pp. 525–536). The Guilford

Press.

Hough, L. M. (2001). I/Owes its advances to personality. In *Personality psychology in the

workplace* (pp. 19–44). American Psychological Association.

https://doi.org/10.1037/10434-001

Hudson, N. W., & Fraley, R. C. (2015). Volitional personality trait change: Can people choose to

change their personality traits? *Journal of Personality and Social Psychology*, *109*(3),

490–507. https://doi.org/10.1037/pspp0000021

Iliev, R., Hoover, J., Dehghani, M., & Axelrod, R. (2016). Linguistic positivity in historical texts

reflects dynamic environmental and psychological factors. *Proceedings of the National

Academy of Sciences*, *113*(49), E7871—-E7879.

https://doi.org/10.1073/pnas.1612058113

IntHout, J., Ioannidis, J. P. A., Borm, G. F., & Goeman, J. J. (2015). Small studies are more

heterogeneous than large ones: A meta-meta-analysis. *Journal of Clinical Epidemiology*,

68(8), 860–869. https://doi.org/10.1016/j.jclinepi.2015.03.017

Jordan, K. N., Sterling, J., Pennebaker, J. W., & Boyd, R. L. (2019). Examining long-term trends

in politics and culture through language of political leaders and cultural institutions.

Proceedings of the National Academy of Sciences, 201811987.

https://doi.org/10.1073/pnas.1811987116

Kahler, C. W., Read, J. P., Wood, M. D., & Palfai, T. P. (2003). Social environmental selection

as a mediator of gender, ethnic, and personality effects on college student drinking.

Psychology of Addictive Behaviors, *17*(3), 226–234. https://doi.org/10.1037/0893-

164X.17.3.226

Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-

system theories. *Perspectives on Psychological Science*, *4*(6), 533–550.

https://doi.org/10.1111/j.1745-6924.2009.01164.x

Kern, M. L., Eichstaedt, J. C., Schwartz, H. A., Dziurzynski, L., Ungar, L. H., Stillwell, D. J.,

Kosinski, M., Ramones, S. M., & Seligman, M. E. P. (2014). The online social self: An

open vocabulary approach to personality. *Assessment*, *21*(2), 158–169.

https://doi.org/10.1177/1073191113514104

Kindon, S., Pain, R., & Kesby, M. (2007). Participatory action research: Origins, approaches,

and methods. In S. Kindon, R. Pain, & M. Kesby (Eds.), *Participatory action research

approaches and methods*. Routledge. https://doi.org/10.4324/9780203933671

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1),

2053951714528481. https://doi.org/10.1177/2053951714528481

Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*(24), 8788–8790. https://doi.org/10.1073/pnas.1320040111

Krasner, L. (1962). Behavior control and social responsibility. *American Psychologist*, *17*(4), 199–204.

Krause, J., James, R., & Croft, D. P. (2010). Personality in the context of social networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1560), 4099–4106. https://doi.org/10.1098/rstb.2010.0216

Krueger, R. F. (2005). Continuity of Axes I and II: Toward a unified model of personality, personality disorders, and clinical disorders. *Journal of Personality Disorders*, *19*(3), 233–261.

Lambiotte, R., & Kosinski, M. (2014). Tracking the digital footprints of personality. *Proceedings of the IEEE*, *102*(12), 1934–1939. https://doi.org/10.1109/JPROC.2014.2359054

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*(4), 475–492. https://doi.org/10.1037/met0000081

Lanning, K., Pauletti, R. E., King, L. A., & McAdams, D. P. (2018). Personality development through natural language. *Nature Human Behaviour*, 1. https://doi.org/10.1038/s41562-018-0329-0

Lazer, D., & Kennedy, R. (2015, October 1). What we can learn from the epic failure of Google Flu Trends. *Wired*. https://www.wired.com/2015/10/can-learn-epic-failure-google-flu-trends/

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. *Science (New York, N.Y.)*, *343*(6176), 1203–1205. https://doi.org/10.1126/science.1248506

Li, Y., Ishi, C. T., Ward, N., Inoue, K., Nakamura, S., Takanashi, K., & Kawahara, T. (2017). Emotion recognition by combining prosody and sentiment analysis for expressing reactive emotion by humanoid robot. *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 1356–1359. https://doi.org/10.1109/APSIPA.2017.8282243

Lien, C.-Y., Bai, G.-J., & Chen, H.-H. (2019). Visited websites may reveal users' demographic information and personality. *2019 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 248–252.

Lin, H.-T., Balcan, M.-F., Hadsell, R., & Ranzato, M. (2020, February 19). Getting started with NeurIPS 2020. *Medium*. https://medium.com/@NeurIPSConf/getting-started-with-neurips-2020-e350f9b39c28

Liu, L., Preotiuc-Pietro, D., Samani, Z. R., Moghaddam, M. E., & Ungar, L. (2016, March 31). Analyzing personality through social media profile picture choice. *Tenth International AAAI Conference on Web and Social Media*. Tenth International AAAI Conference on Web and Social Media. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13102

Maltby, J., Wood, A. M., Day, L., & Pinto, D. (2012). The position of authenticity within extant models of personality. *Personality and Individual Differences*, *52*(3), 269–273. https://doi.org/10.1016/j.paid.2011.10.014

Manzo, L. C., & Brightbill, N. (2007). Toward a participatory ethics. In S. Kindon, R. Pain, &

M. Kesby (Eds.), *Participatory action research approaches and methods*. Routledge.

https://doi.org/10.4324/9780203933671

Marks, D. F. (2018). American psychologists, the Central Intelligence Agency, and enhanced

interrogation. *Health Psychology Open*, *5*(2). https://doi.org/10.1177/2055102918796610

Martin, J. (2019, April 10). What the bible says about secrets. *The New York Times*.

https://www.nytimes.com/2019/04/10/opinion/sunday/privacy-religion.html

Matz, S. C., Gladstone, J. J., & Stillwell, D. J. (2016). Money buys happiness when spending fits

our personality. *Psychological Science*, *27*(5), 715–725.

https://doi.org/10.1177/0956797616635200

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an

effective approach to digital mass persuasion. *Proceedings of the National Academy of

Sciences*, *114*(48), 12714–12719. https://doi.org/10.1073/pnas.1710966114

Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition & Emotion*,

23(2), 209–237. https://doi.org/10.1080/02699930802204677

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication

crisis? What does "failure to replicate" really mean? *American Psychologist*, *70*(6), 487–

498. https://doi.org/10.1037/a0039400

Mayer, J. D. (2005). A tale of two visions: Can a new view of personality help integrate

psychology? *American Psychologist*, *60*(4), 294–307. https://doi.org/10.1037/0003-

066X.60.4.294

Mayer, J. D. (2007). The big questions of Personality Psychology: Defining common pursuits of

the discipline. *Imagination, Cognition and Personality*. https://doi.org/10.2190/IC.27.1.b

Mayer, J. D., & Allen, J. L. (2013). A personality framework for the unification of Psychology. *Review of General Psychology*, *17*(2), 196–202. https://doi.org/10.1037/a0032934

McAbee, S. T., & Connelly, B. S. (2016). A multi-rater framework for studying personality: The trait-reputation-identity model. *Psychological Review*, *123*(5), 569–591. https://doi.org/10.1037/rev0000035

McMullan, T. (2015, July 23). What does the panopticon mean in the age of digital surveillance? *The Guardian*. https://www.theguardian.com/technology/2015/jul/23/panopticon-digital-surveillance-jeremy-bentham

Meehl, P. E. (1945). The dynamics of "structured" personality tests. *Journal of Clinical Psychology*, *1*(4), 296–303. https://doi.org/10.1002/1097-4679(194510)1:4<296::AID-JCLP2270010410>3.0.CO;2-#

Meerloo, J. A. M. (1956). *The rape of the mind: The psychology of thought control, menticide, and brainwashing*. World Pub. Co.

Mehl, M. R. (2017). The Electronically Activated Recorder (EAR): A method for the naturalistic observation of daily social behavior. *Current Directions in Psychological Science*, *26*(2), 184–190. https://doi.org/10.1177/0963721416680611

Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science*, *7*(3), 221–237. https://doi.org/10.1177/1745691612441215

Mitchell, L., Frank, M. R., Harris, K. D., Dodds, P. S., & Danforth, C. M. (2013). The geography of happiness: Connecting Twitter sentiment and expression, demographics, and objective characteristics of place. *PLOS ONE*, *8*(5), e64417. https://doi.org/10.1371/journal.pone.0064417

Mitra, T., Counts, S., & Pennebaker, J. W. (2016). Understanding anti-vaccination attitudes in

social media. *Proceedings of the Tenth International AAAI Conference on Web and

Social Media*. International AAAI Conference on Web and Social Media.

https://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13073

Mozur, P. (2018, July 8). Inside China's dystopian dreams: A.I., shame and lots of cameras. *The

New York Times*. https://www.nytimes.com/2018/07/08/business/china-surveillance-

technology.html

Murray, H. A. (1938). *Explorations in personality*. Oxford Univ. Press.

Narayanan, A., & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. *2008

IEEE Symposium on Security and Privacy (Sp 2008)*, 111–125.

https://doi.org/10.1109/SP.2008.33

Negley, G. (1966). Philosophical views on the value of privacy. *Law and Contemporary

Problems*, *31*(2), 319–325.

Nye, C. D., Allemand, M., Gosling, S. D., Potter, J., & Roberts, B. W. (2016). Personality trait

differences between young and middle-aged adults: Measurement artifacts or actual

trends? *Journal of Personality*, *84*(4), 473–492. https://doi.org/10.1111/jopy.12173

Oboler, A., Welsh, K., & Cruz, L. (2012). The danger of big data: Social media as computational

social science. *First Monday*, *17*(7). https://doi.org/10.5210/fm.v17i7.3993

Obschonka, M., Stuetzer, M., Rentfrow, P. J., Shaw-Taylor, L., Satchell, M., Silbereisen, R. K.,

Potter, J., & Gosling, S. D. (2018). In the shadow of coal: How large-scale industries

contributed to present-day regional differences in personality and well-being. *Journal of

Personality and Social Psychology*, *115*(5), 903–927.

https://doi.org/10.1037/pspp0000175

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Penguin Random House USA Ex.

O'Neil, C., & Schutt, R. (2013). *Doing data science: Straight talk from the frontline* (1 edition). O'Reilly Media.

Orwell, G. (1949). *Nineteen eighty-four: A novel*. Secker and Warburg.

Pan, W., Altshuler, Y., & Pentland, A. (Sandy). (2012). Decoding social influence and the wisdom of the crowd in financial trading network. *MIT Web Domain*. https://dspace.mit.edu/handle/1721.1/80764

Pardo, Y., Aguilar, R., Molinuevo, B., & Torrubia, R. (2007). Alcohol use as a behavioural sign of disinhibition: Evidence from J.A. Gray's model of personality. *Addictive Behaviors*, *32*(10), 2398–2403. https://doi.org/10.1016/j.addbeh.2007.02.010

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., Ungar, L. H., & Seligman, M. E. P. (2015). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, *108*(6), 934–952. https://doi.org/10.1037/pspp0000020

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In *Handbook of research methods in personality psychology* (pp. 224–239). Guilford.

Paxton, A., & Griffiths, T. L. (2017). Finding the traces of behavioral and cognitive processes in big data and naturally occurring datasets. *Behavior Research Methods*, *49*(5), 1630–1638. https://doi.org/10.3758/s13428-017-0874-x

Pennebaker, J. W., Chung, C. K., Frazee, J., Lavergne, G. M., & Beaver, D. I. (2014). When small words foretell academic success: The case of college admissions essays. *PLOS ONE*, *9*(12), e115844.

Powell, R. A., Digdon, N., Harris, B., & Smithson, C. (2014). Correcting the record on Watson, Rayner, and Little Albert: Albert Barger as "psychology's lost boy." *The American Psychologist*, *69*(6), 600–611. https://doi.org/10.1037/a0036854

Power, C., & Elliott, J. (2006). Cohort profile: 1958 British birth cohort (National Child Development Study). *International Journal of Epidemiology*, *35*(1), 34–41. https://doi.org/10.1093/ije/dyi183

Prilleltensky, I. (1989). Psychology and the status quo. *American Psychologist*, *44*(5), 795–802. https://doi.org/10.1037/0003-066X.44.5.795

Prince, S. A., Adamo, K. B., Hamel, M. E., Hardt, J., Gorber, S. C., & Tremblay, M. (2008). A comparison of direct versus self-report measures for assessing physical activity in adults: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity*, *5*(1), 56. https://doi.org/10.1186/1479-5868-5-56

Pryor, J. B., Gibbons, F. X., Wicklund, R. A., Fazio, R. H., & Hood, R. (1977). Self-focused attention and self-report validity. *Journal of Personality*, *45*(4), 513–527. https://doi.org/10.1111/j.1467-6494.1977.tb00168.x

Quodling, A. (2018, April 13). Shadow profiles—Facebook knows about you, even if you're not on Facebook. *The Conversation*. https://theconversation.com/shadow-profiles-facebook-knows-about-you-even-if-youre-not-on-facebook-94804

Rauthmann, J. F. (2017). Development of implicit personality. In J. Specht (Ed.), *Personality Development Across the Lifespan* (pp. 537–554). Academic Press. https://doi.org/10.1016/B978-0-12-804674-6.00032-6

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., Ziegler, M., Jones, A. B., & Funder, D. C. (2014). The Situational Eight DIAMONDS: A

taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, *107*(4), 677–718. https://doi.org/10.1037/a0037250

Rauthmann, J. F., & Sherman, R. A. (2018). The description of situations: Towards replicable domains of psychological situation characteristics. *Journal of Personality and Social Psychology*, *114*(3), 482–488. https://doi.org/10.1037/pspp0000162

Reece, A. G., & Danforth, C. M. (2017). Instagram photos reveal predictive markers of depression. *EPJ Data Science*, *6*(1), 1–12. https://doi.org/10.1140/epjds/s13688-017-0110-z

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), Pub. L. No. 32016R0679, 119 OJ L (2016). http://data.europa.eu/eli/reg/2016/679/oj/eng

Reiss, J., & Sprenger, J. (2017). Scientific objectivity. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017). Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2017/entries/scientific-objectivity/

Rexroth, K. (1959). *Bird in the bush: Obvious essays*. New Directions.

Reynolds, B., Ortengren, A., Richards, J. B., & Wit, H. de. (2006). Dimensions of impulsive behavior: Personality and behavioral measures. *Personality and Individual Differences*, *40*(2), 305–315. https://doi.org/10.1016/j.paid.2005.03.024

Reynolds, K. J., Turner, J. C., Branscombe, N. R., Mavor, K. I., Bizumic, B., & Subašić, E. (2010). Interactionism in personality and social psychology: An integrated approach to

understanding the mind and behaviour. *European Journal of Personality*, *24*(5), 458–482. https://doi.org/10.1002/per.782

Roberts, J. J. (2018, May 24). *The GDPR is in effect: A plain English guide for US companies*. Fortune. https://fortune.com/2018/05/24/the-gdpr-is-in-effect-should-u-s-companies-be-afraid/

Rosenthal, S. A., & Hooley, J. M. (2010). Narcissism assessment in social–personality research: Does the association between narcissism and psychological health result from a confound with self-esteem? *Journal of Research in Personality*, *44*(4), 453–465. https://doi.org/10.1016/j.jrp.2010.05.008

Roski, J., Bo-Linn, G. W., & Andrews, T. A. (2014). Creating value in health care through big data: Opportunities and policy implications. *Health Affairs*, *33*(7), 1115–1122. https://doi.org/10.1377/hlthaff.2014.0147

Sariyska, R., & Montag, C. (2019). An overview on doing psychodiagnostics in Personality Psychology and tracking physical activity via smartphones. In H. Baumeister & C. Montag (Eds.), *Digital Phenotyping and Mobile Sensing: New Developments in Psychoinformatics* (pp. 45–63). Springer International Publishing. https://doi.org/10.1007/978-3-030-31620-4_4

Saxbe, D. E., Yang, X.-F., Borofsky, L. A., & Immordino-Yang, M. H. (2013). The embodiment of emotion: Language use during the feeling of social emotions predicts cortical somatosensory activity. *Social Cognitive and Affective Neuroscience*, *8*(7), 806–812. https://doi.org/10.1093/scan/nss075

Schiele, B. C., Baker, A. B., & Hathaway, S. R. (1943). The Minnesota Multiphasic Personality Inventory. *Journal-Lancet*, *63*, 292–297.

Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology*, *80*(5), 607–620. https://doi.org/10.1037/0021-9010.80.5.607

Segalin, C., Celli, F., Polonio, L., Kosinski, M., Stillwell, D., Sebe, N., Cristani, M., & Lepri, B. (2017). What your Facebook profile picture reveals about your personality. *Proceedings of the 25th ACM International Conference on Multimedia*, 460–468. https://doi.org/10.1145/3123266.3123331

Seligman, M. (2018). The Hoffman report, the Central Intelligence Agency, and the defense of the nation: A personal view. *Health Psychology Open*, *5*(2). https://doi.org/10.1177/2055102918796192

Settanni, M., Azucar, D., & Marengo, D. (2018). Predicting individual characteristics from digital traces on social media: A meta-analysis. *Cyberpsychology, Behavior, and Social Networking*, *21*(4), 217–228. https://doi.org/10.1089/cyber.2017.0384

Sharma, L., Markon, K. E., & Clark, L. A. (2014). Toward a theory of distinct types of "impulsive" behaviors: A meta-analysis of self-report and behavioral measures. *Psychological Bulletin*, *140*(2), 374–408. https://doi.org/10.1037/a0034418

Sherman, R. A., Nave, C. S., & Funder, D. C. (2010). Situational similarity and personality predict behavioral consistency. *Journal of Personality and Social Psychology*, *99*(2), 330–343. https://doi.org/10.1037/a0019796

Shiramizu, V. K. M., Kozma, L., DeBruine, L. M., & Jones, B. C. (2019). Are dark triad cues really visible in faces? *Personality and Individual Differences*, *139*, 214–216. https://doi.org/10.1016/j.paid.2018.11.011

Shobeiri, S., Laroche, M., & Mazaheri, E. (2013). Shaping e-retailer's website personality: The importance of experiential marketing. *Journal of Retailing and Consumer Services*, *20*(1), 102–110. https://doi.org/10.1016/j.jretconser.2012.10.011

Shokri, R., Theodorakopoulos, G., Le Boudec, J.-Y., & Hubaux, J.-P. (2011). Quantifying location privacy. *2011 IEEE Symposium on Security and Privacy*, 247–262. https://doi.org/10.1109/SP.2011.18

Soto, C. J. (2019). How replicable are links between personality traits and consequential life outcomes? The life outcomes of personality replication project. *Psychological Science*, *30*(5), 711–727. https://doi.org/10.1177/0956797619831612

Staiano, J., Lepri, B., Aharony, N., Pianesi, F., Sebe, N., & Pentland, A. (2012). Friends don't lie: Inferring personality traits from social network structure. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 321–330. https://doi.org/10.1145/2370216.2370266

Stanton, A. M., Boyd, R. L., Pulverman, C. S., & Meston, C. M. (2015). Determining women's sexual self-schemas through advanced computerized text analysis. *Child Abuse and Neglect*, *46*. https://doi.org/10.1016/j.chiabu.2015.06.003

Stanton, A. M., Meston, C. M., & Boyd, R. L. (2017). Sexual self-schemas in the real world: Investigating the ecological validity of language-based markers of childhood sexual abuse. *Cyberpsychology, Behavior, and Social Networking*, *20*(6). https://doi.org/10.1089/cyber.2016.0657

Stella, M., Ferrara, E., & Domenico, M. D. (2018). Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences*, 201803470. https://doi.org/10.1073/pnas.1803470115

Strong, E. K. (1934). The Vocational Interest Test. *The Vocational Guidance Magazine*, *12*(8), 49–56. https://doi.org/10.1002/j.2164-5884.1934.tb00594.x

Strong, E. K. (1935). Predictive value of the vocational interest test. *Journal of Educational Psychology*, *26*(5), 331–349. https://doi.org/10.1037/h0062498

Sugiura, L., Wiles, R., & Pope, C. (2017). Ethical challenges in online research: Public/private perceptions. *Research Ethics*, *13*(3–4), 184–199. https://doi.org/10.1177/1747016116650720

Sun, J., Schwartz, H. A., Son, Y., Kern, M. L., & Vazire, S. (2019). The language of well-being: Tracking fluctuations in emotion experience through everyday speech. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/pspp0000244

Sweeney, L. (2005, June 15). *Privacy technologies for homeland security. Statement before the Privacy and Integrity Advisory Committee of the Department of Homeland Security*.

Tellakat, M., Boyd, R. L., & Pennebaker, J. W. (2019). How do online learners study? The psychometrics of students' clicking patterns in online courses. *PLOS ONE*, *14*(3), 1–17. https://doi.org/10.1371/journal.pone.0213863

Thompson, S. A., & Warzel, C. (2019, December 19). Twelve million phones, one dataset, zero privacy. *The New York Times*. https://www.nytimes.com/interactive/2019/12/19/opinion/location-tracking-cell-phone.html

Thorstad, R., & Wolff, P. (2019). Predicting future mental illness from social media: A big-data approach. *Behavior Research Methods*, *51*(4), 1586–1600. https://doi.org/10.3758/s13428-019-01235-z

Tkaczynski, P. J., Ross, C., MacLarnon, A., Mouna, M., Majolo, B., & Lehmann, J. (2018).

Measuring personality in the field: An In Situ comparison of personality quantification

methods in wild Barbary macaques (Macaca sylvanus). *Journal of Comparative

Psychology*, No Pagination Specified-No Pagination Specified.

https://doi.org/10.1037/com0000163

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence

from faces predict election outcomes. *Science (New York, N.Y.)*, *308*(5728), 1623–1626.

https://doi.org/10.1126/science.1110589

Torra, V. (2017). *Data privacy: Foundations, new developments and the big data challenge*.

Springer.

Tufekci, Z. (2019, April 21). Think you're discreet online? Think again. *The New York Times*.

https://www.nytimes.com/2019/04/21/opinion/computational-inference.html

Tunick, M. (2014). *Balancing privacy and free speech: Unwanted attention in the age of social

media* (1 edition). Routledge.

Turkyilmaz, C. A., Erdem, S., & Uslu, A. (2015). The effects of personality traits and website

quality on online impulse buying. *Procedia - Social and Behavioral Sciences*, *175*, 98–

105. https://doi.org/10.1016/j.sbspro.2015.01.1179

Vallacher, R. R., Read, S. J., & Nowak, A. (Eds.). (2017). *Computational social psychology*.

Routledge.

Walker, M., Schönborn, S., Greifeneder, R., & Vetter, T. (2018). The Basel Face Database: A

validated set of photographs reflecting systematic differences in Big Two and Big Five

personality dimensions. *PLoS ONE*, *13*(3).

Warren, S. D., & Brandeis, L. D. (1890). The right to privacy. *Harvard Law Review*, *4*(5), 193–
220. JSTOR. https://doi.org/10.2307/1321160

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*(2), 158–
177. https://doi.org/10.1037/h0074428

Wilson, G. D. (1990). Personality, time of day and arousal. *Personality and Individual
Differences*, *11*(2), 153–168. https://doi.org/10.1016/0191-8869(90)90008-F

Wilt, J., Noftle, E. E., Fleeson, W., & Spain, J. S. (2011). The dynamic role of personality states
in mediating the relationship between extraversion and positive affect. *Journal of
Personality*, *80*(5), 1205–1236. https://doi.org/10.1111/j.1467-6494.2011.00756.x

Wojcik, S. P., Hovasapian, A., Graham, J., Motyl, M., & Ditto, P. H. (2015). Conservatives
report, but liberals display, greater happiness. *Science*, *347*(6227), 1243–1246.
https://doi.org/10.1126/science.1260817

Wrzus, C., & Mehl, M. R. (2015). Lab and/or field? Measuring personality processes and their
social consequences. *European Journal of Personality*, *29*(2), 250–271.
https://doi.org/10.1002/per.1986

Wu, T. (2019, April 10). How capitalism betrayed privacy. *The New York Times*.
https://www.nytimes.com/2019/04/10/opinion/sunday/privacy-capitalism.html

Yin, S., & Kaynak, O. (2015). Big data for modern industry: Challenges and trends. *Proceedings
of the IEEE*, *103*(2), 143–146. https://doi.org/10.1109/JPROC.2015.2388958

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*(4), 1036–1040. https://doi.org/10.1073/pnas.1418680112

Yuan, Y. (2020, February 24). YOLO creator Joseph Redmon stopped CV research due to

ethical concerns. *Medium*. https://medium.com/syncedreview/yolo-creator-says-he-

stopped-cv-research-due-to-ethical-concerns-b55a291ebb29

Zimbardo, P. G. (1973). On the ethics of intervention in human psychological research: With

special reference to the Stanford prison experiment. *Cognition*, *2*(2), 243–256.

https://doi.org/10.1016/0010-0277(72)90014-5

Zimmer, M. (2018). Addressing conceptual gaps in big data research ethics: An application of

contextual integrity. *Social Media + Society*, *4*(2), 2056305118768300.

https://doi.org/10.1177/2056305118768300