Personality Sensing for Theory Development and Assessment in the Digital Age

GABRIELLA M. HARARI[1]*, SUMER S. VAID[1], SANDRINE R. MÜLLER[2], CLEMENS STACHL[1],
ZACHARIAH MARRERO[3], RAMONA SCHOEDEL[4], MARKUS BÜHNER[4] and SAMUEL D. GOSLING[3,5]

[1]*Department of Communication, Stanford University, Stanford, CA USA*
[2]*Data Science Institute, Columbia University, New York City, NY USA*
[3]*Department of Psychology, University of Texas at Austin, Austin, TX USA*
[4]*Department of Psychology, Ludwig Maximilian University of Munich, Munich, Germany*
[5]*Melbourne School of Psychological Sciences, University of Melbourne, Melbourne, Australia*

Abstract: People around the world own digital media devices that mediate and are in close proximity to their daily behaviours and situational contexts. These devices can be harnessed as sensing technologies to collect information from sensor and metadata logs that provide fine-grained records of everyday personality expression. In this paper, we present a conceptual framework and empirical illustration for personality sensing research, which leverages sensing technologies for personality theory development and assessment. To further empirical knowledge about the degree to which personality-relevant information is revealed via such data, we outline an agenda for three research domains that focus on the description, explanation, and prediction of personality. To illustrate the value of the personality sensing research agenda, we present findings from a large smartphone-based sensing study (N = 633) characterizing individual differences in sensed behavioural patterns (physical activity, social behaviour, and smartphone use) and mapping sensed behaviours to the Big Five dimensions. For example, the findings show associations between behavioural tendencies and personality traits and daily behaviours and personality states. We conclude with a discussion of best practices and provide our outlook on how personality sensing will transform our understanding of personality and the way we conduct assessment in the years to come. © 2020 European Association of Personality Psychology

Key words: personality; behaviour; situational contexts; mobile sensing; naturalistic observation

INTRODUCTION

Around 300 BC, Theophrastus published his great work, *Characters*, in which he described 30 types of Athenian people (e.g. the chatty man and the gossip); he illustrated his character portraits with detailed descriptions of how these different types of people typically behaved. Theophrastus' work rested on two key points that remain relevant to personality science today. First, taxonomic classification can be used to summarize patterns in people's characteristic behaviours via personality types and traits. Second, these types and traits have implications for a wide range of important outcomes.

In the two millennia since Theophrastus wrote his book, the technology for gathering behavioural information has essentially remained unchanged. Researchers still generally rely on explicit reports by the self and observers to record what people do (Baumeister, Vohs, & Funder, 2007; Funder, 2009). But these methods are subject to a wide range of limitations (e.g. response biases and memory limitations; Paulhus & Vazire, 2007) that potentially undermine their validity and the viability of their widespread use for understanding personality expression in daily life. However, over the past decade, advances in mobile sensing, along with the large-scale adoption of sensing technologies (e.g. smartphones, wearables, and home appliances), have paved the way to realize the vision first set out by Theophrastus 2000 years ago by providing detailed descriptions of how people behave.

Sensing technologies have made it possible to use digital media devices to collect data about people's thoughts, feelings, and behaviours in an ecologically sensitive manner, longitudinally, across situational contexts, and at scale (Campbell et al., 2008; Eagle & Pentland, 2003; Harari et al., 2016; Lane, Miluzzo, Lu, Peebles, & Choudhury, 2010; Lathia et al., 2013; Miller, 2012; Wrzus & Mehl, 2015). So here, we propose an agenda for *personality sensing* research, which aims to leverage sensing technologies for personality theory development and assessment. We believe that personality sensing research should have at least two main goals: (i) To address basic theoretical questions about the structure and dynamics of personality expression as it plays out in everyday life; and (ii) to develop automated personality and life outcome assessment techniques by creating models derived from sensing data about everyday behaviour and situational contexts.

*Correspondence to: Gabriella M. Harari, Department of Communication, Stanford University, 450 Jane Stanford Way, Stanford, CA 94305, USA.
E-mail: gharari@stanford.edu*

By combining theoretical and methodological approaches from the fields of psychology and computer science, personality sensing is poised to advance the field's understanding of the behavioural and contextual aspects of personality. Doing so requires an intensive 'big data' approach to collecting personality-relevant information that offers a high-fidelity picture of personality expression in everyday life. To meet the goals of personality sensing research, the data must be broad-based (many kinds of behavioural data), fine-grained (many behavioural observations across contexts), longitudinal (many days, weeks, or even years of behavioural data), and representative (behavioural data from many participants across diverse samples). Such an approach to understanding and assessing personality will provide psychologists with the tools necessary to address long-standing questions about the behavioural patterns of day-to-day life that predict consequential life outcomes (Ozer & Benet-Martínez, 2005; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007; Soto, 2019).

In this paper, we first define personality sensing research and distinguish the different sensing technologies and data sources (i.e. mobile sensors and metadata) that reveal personality-relevant information. We provide a brief summary of sensing research relevant to personality, focusing on information about behavioural, person, and contextual factors. Next, we outline a research agenda for guiding future personality sensing research that focuses on three empirical research domains—description, explanation, and prediction. To illustrate the value of our proposed agenda, we present exploratory findings from our own smartphone-based personality sensing study as an empirical example that maps sensed behaviours of young adults onto the Big Five personality dimensions. In particular, we examine three behavioural domains (physical activity, social interactions, and mobile phone use) to describe and explain individual differences in behavioural tendencies and dynamics. Finally, we conclude with our outlook on some of the core theoretical, methodological, and ethical considerations for personality sensing research.

PERSONALITY SENSING: DEFINITION, SCOPE, AND MEASUREMENT

Personality sensing can be described as the collection and modelling of sensing data for the understanding and assessment of personality expression in daily life. Sensing data can be collected from ubiquitous computing devices such as smartphones, wearable devices, smart home appliances, and smart cars, which can have sensing software installed on them to permit the collection of mobile sensor data and metadata logs. These sources of 'raw' sensing data (i.e. unprocessed device data) can then be processed and aggregated to obtain personality-relevant information about a person's thoughts and feelings, behaviours, and situational contexts (Funder, 2001, 2006).

Figure 1 presents a schematic overviewing the different levels of aggregation and types of personality-relevant information that can be obtained from sensing data. Sensing data generally involves the collection of raw sensor data (e.g. latitude and longitude coordinate data from global positioning system [GPS] sensors) and/or metadata logs (e.g. time-stamped call logs) that need to be pre-processed to obtain specific behavioural and situational 'features' or variables (e.g. number of incoming and outgoing calls derived from call logs or places visited derived from GPS data). These features are sensor-based indicators of broader behavioural domains that can provide personality-relevant information about a person's characteristic patterns (e.g. stable tendencies and dynamic patterns). For example, number of calls and places visited are narrow indicators of the broader domains of social behaviour and situational cues, which capture a person's tendency to socially interact with others and spend time in different physical environments.

Typically, sensing features tend to represent objective indicators of behaviours and situational contexts. However, such objective behavioural and situational indicators often do not reflect psychologically active information about how people are perceiving the world around them (e.g. their motivations for engaging in a behaviour or being in a situational context). To obtain sensed information that is psychologically richer in reflecting people's subjective psychological experiences, language data must be collected from people's voices (via the microphone) or the words they type into the phone (via text message content, keystroke logging, or screenshots of smartphone screen content; see section on *Person Factors* for more information). Such sources of sensing data can also be used to obtain features that characterize behaviours and situations in more depth (e.g. the content or topic of
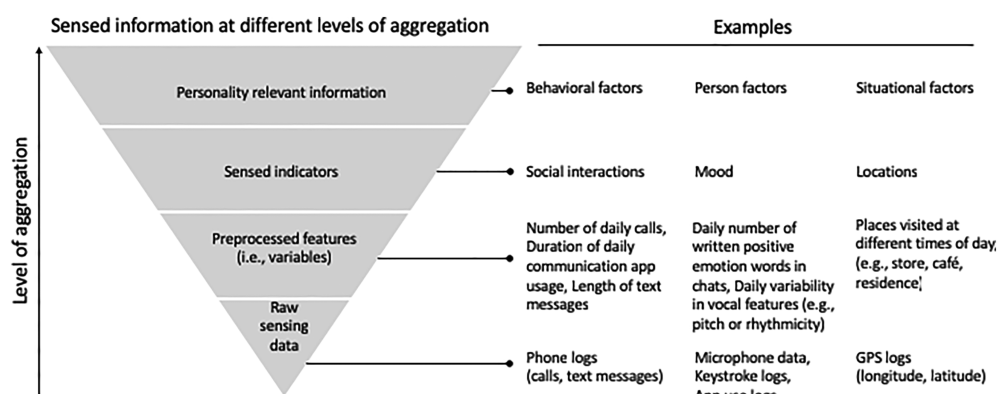

FIGURE 1.   Schematic overview of personality-relevant information from sensing data.

conversations and types of interaction partners), going beyond estimates of degree (e.g. frequency and/or duration) to capture more qualitative information about the observations being made. Alternatively, self-report methods (e.g. experience sampling and daily diaries) can be used as a complementary method to obtain information about subjective experiences by asking people to report on them directly.

Of the various sensing technologies, smartphones in particular stand out as being one of the most promising sources of personality-relevant information (Harari, Müller, Aung, & Rentfrow, 2017; Harari, Müller, & Gosling, 2018). Therefore, we draw heavily on smartphones in our description of the personality sensing agenda and in our empirical illustration on the types of personality-relevant information that can be obtained from sensors. But in principle, many of the research opportunities and illustrations presented here could also be pursued with other sensing technologies (e.g. wearables, smart home devices, and smart cars) because these devices all share common data sources via their sensors and metadata logs (e.g. accelerometers, microphones, and device usage logs). In the coming years, we anticipate personality sensing with smartphones and other devices will become part of the standard research toolkit in moving towards a truly behavioural personality science (Buss & Craik, 1983; Furr, 2009).

PERSONALITY-RELEVANT INFORMATION FROM SENSING DATA

The personality triad framework dictates that to understand personality expression, we must account for three sources of personality-relevant information (Funder, 2006): behavioural factors, person factors, and situational factors. Behavioural factors refer to objectively observable behavioural acts (i.e. what people do). Person factors refer to people's characteristic individual differences and subjective experiences (i.e. what people think and feel). Situational factors refer to contextual information that may be influencing momentary personality expression (i.e. what environments people are in). Thus, we define personality sensing as the use of sensing technologies for understanding and assessing behaviours, persons, and situations in daily life. In line with this personality triad approach, we distinguish three different types of personality-relevant information that can be captured from sensing data.

Behavioural factors: Objective assessment of everyday behaviours

Behaviour can be operationalized as 'verbal utterances (excluding verbal reports in psychological assessment contexts) or movements that are potentially available to careful observers using normal sensory processes' (Furr, 2009; p. 372). By this definition, there are many different types of behaviour that psychologists may find interesting to understand and assess. In the context of personality sensing research, the assessment of behaviours that reflect a person's lifestyle is the most common, and direct, way of using

sensing data. This approach to behavioural assessment uses the sensing technology as an observational tool that mediates or is in close proximity to many kinds of everyday behaviours that can be measured via mobile sensors and metadata logs (for reviews of sensing data in smartphones, see Lane et al., 2010; Harari et al., 2016). Three broad domains of behavioural information that can be directly collected or indirectly inferred from smartphone sensing data include (i) movement behaviours (e.g. physical activity and mobility patterns), (ii) social interactions (e.g. in-person and mediated communications), and (iii) daily life activities (e.g. mediated and non-mediated activities; Harari et al., 2017). For example, the accelerometer, microphone, and light sensor can be used to collect data about peoples' activity levels (e.g. walking and running; Miluzzo, Papandrea, Lane, Lu, & Campbell, 2010) surrounding acoustic environment (e.g. listening to music and having a conversation; Lu, Pan, Lane, & Choudhury, 2009), and sleeping patterns (e.g. duration and sleep and wake-up times; Chen et al., 2013). The phone logs can be used to collect data about peoples' social interactions (e.g. social media app use and incoming and outgoing calls and text messages; Chittaranjan, Blom, & Gatica-Perez, 2013; Harari et al., 2020) and general application use (e.g. gaming apps and photography apps; Schoedel et al., 2019; Stachl et al., 2017).

Some of the current limitations of sensing for behavioural assessment stem from the trade-off between capturing the maximum amount of information that can be sensed and adopting more privacy-preserving aggregation strategies (for a discussion, see Beierle et al., 2018). For example, to understand and assess social behaviour one might collect raw audio data from the microphone sensor (containing the content of conversations; e.g. Mehl, 2017) or collect on-device classifications of whether a person is around voices or engaged in conversation (Choudhury & Basu, 2005; Schmid Mast, Gatica-Perez, Frauendorfer, Nguyen, & Choudhury, 2015). Such privacy-preserving approaches tend to rely on classifier models (which need to be well-validated and updated over time) and/or rely on some form of pre-processing on the device to aggregate and collect frequency and duration estimates of behaviour that are reminiscent of theoretical perspectives like the Act Frequency approach (and may suffer from similar limitations; Block, 1989; Buss & Craik, 1983).

Person factors: Capturing individual differences in thoughts and feelings

A person's thoughts and feelings can be assessed using sensing technologies via two mechanisms: (i) collecting passively sensed inferences from mobile sensors and metadata and/or (ii) collecting traditional subjective self-reports provided by participants via the sensing device. The passive sensing approach involves collecting and processing sensing data that can be used to make inferences about an individual's cognitive and affective states via sensors and metadata logs as well as the words a person uses, which are thought to reflect personality-relevant information about people's thoughts and feelings (Boyd & Pennebaker, 2017). For

example, language information can be directly collected via voice data from microphones (Mehl, 2017), keystrokes logged by custom keyboards (Buschek, Bisinger, & Alt, 2018), or extracted from screenshots of smartphone content (Ram et al., 2020; Reeves et al., 2019). Early studies sought to show that passive sensing is a promising approach to assessing cognitive and affective states from sensors and metadata logs by modelling stress from vocal features in microphone data (Lu et al., 2012), alertness and attention from phone usage features and app-use logs, respectively (Abdullah et al., 2016; Murnane et al., 2016), and mood from phone usage features (Likamwa, Liu, Lane, & Zhong, 2011). However, many of the early studies were conducted with small samples ($N < 50$), so more research is needed (e.g. to replicate past study findings with larger samples, validate prediction models developed) before passive assessment of thoughts and feelings from sensing data is considered a reliable and valid approach.

In contrast, the self-report approach involves querying people about their subjective thoughts and feelings (e.g. by sending them notifications to respond to experience sampling surveys). This approach uses the sensing device as an ambulatory assessment tool for delivering survey questions, allowing individuals to provide self-reports of their psychological states that can be used to complement and enrich the passively sensed data with information about people's subjective experiences (e.g. to understand the relationships between happiness and physical activity and mood and physical context; Lathia, Sandstrom, Mascolo, & Rentfrow, 2017; Muller, Peters, Matz, Wang, & Harari, in press; Sandstrom, Lathia, Mascolo, & Rentfrow, 2017).

Situational factors: Inferring contextual information

Sensing technologies permit recording of information about a person's social context and physical environment (Abowd & Dey, 1999). In particular, situational cues can be assessed using sensing data to obtain objective contextual information (e.g. the who, where, and when of a situation; Rauthmann, Sherman, & Funder, 2015). However, the assessment of contextual information that reflects the psychological construal of a given situation currently requires the use of traditional self-reporting methods as a complementary approach to passive sensing to obtain people's subjective perceptions of the situation (e.g. Müller et al., 2017; Vaizman, Ellis, & Lanckriet, 2017;).

The passive sensing approach to situational assessment relies on both direct observation and inferred contextual information. For example, GPS data directly reveals the geographical location of an individual via the longitude and latitude coordinates. When processed, GPS coordinates also reveal more semantic place labels (e.g. that a person is in a residence, store, or cafe). Moreover, combinations of various sensing data sources can reveal more nuanced situational information (e.g. combining GPS data with microphone data to infer that a person is likely at home and alone; see Harari et al., 2018 for an overview of smartphone sensing for situational assessment). Finally, sensing data can be combined with other external data sources, such as publicly available regional data (e.g. crime statistics and census data), to further enrich the contextual information available (e.g. obtaining indexes of safety and population characteristics). For example, GPS data can be enriched with elevation data via the Google application programming interface to get a sense of the topographic features of the environment (Stachl et al., 2019). Taken together, studies that adopt the sensing approach to assessing situational factors have shown that sensing technologies are a promising approach to automating the collection of contextual information in daily life.

AN AGENDA FOR PERSONALITY SENSING RESEARCH

Having introduced a conceptual framework of personality-relevant information from sensing data, we now outline three domains of personality sensing research (see Figure 2). For conceptual clarity, we describe these domains in the following sequence: description, explanation, and prediction (e.g. Shmueli, 2011; Yarkoni & Westfall, 2017); however, we do not suggest that these domains must follow each other in a given sequence. In fact, one of the unique features of personality sensing approaches is that one might be interested in making predictions (e.g. of an important personality-relevant outcome like occupational satisfaction), without necessarily needing to describe or explain the types of personality-relevant sensing information that lead to the prediction.

Describing personality expression

Sensing technologies are well suited for the description of real-world personality expression in daily life, continuously, and with fine granularity. Over the past two decades, researchers have called for a better descriptive understanding of psychological phenomena, including everyday thoughts and feelings (Boyd & Pennebaker, 2017; Fast & Funder, 2008), behaviours (Baumeister et al., 2007; Funder, 2006; Rozin, 2001; Zeigler-Hill, Shackelford, Nave, Feeney, & Furr, 2018), and situations (Rauthmann et al., 2015; Rauthmann & Sherman, 2018). One particularly promising approach for personality sensing research describing real-world personality expressions is to identify new traits and types derived from sensing data that reflect people's everyday behavioural and contextual patterns. Such an approach would provide a descriptive understanding of individual differences in behavioural and contextual patterns that may reliably distinguish individuals from one another and predict consequential life outcomes (e.g. in the domains of health, work, and relationships). Of course, this approach would yield only the traits and types related to the kinds of behaviours and situational contexts that can be sensed from these technologies. So, if one's ultimate goal is to inform personality theory (vs. maximize the predictive ability of models designed to assess personality or life outcomes), some care must be taken in deriving theoretically meaningful and human-interpretable features for use in the analyses. For example, researchers might consider whether the features

FIGURE 2. Examples of research questions and analyses for three empirical domains of personality sensing research.

reflect or can be explained by tendencies in how people tend to think, feel, and behave in the world.

To understand how behaviours and contexts are expressed in daily life, large-scale longitudinal studies that collect sensing data are needed. Such studies will permit researchers to examine the stability and change associated with everyday behaviours and contexts, such as movement behaviours (e.g. physical activity and mobility; Harari, Müller, Mishra, et al., 2017; Muller et al., in press), social behaviours (e.g. conversations, calling, texting, and app use; Harari et al., 2020; Montag et al., 2015; Stachl et al., 2017), everyday activities (e.g. sleeping and using one's phone; Schoedel et al., 2020 this issue), and situations (e.g. places visited and ambience; Santani et al., 2016). For example, past work on sociability tendencies of young adults described rates of conversation, calling, texting, and mobile app use and showed that such behaviours have high degrees of between-person variability and consistency from day to day (Harari et al., 2020). A descriptive understanding of persons, behaviours, and situations as they occur in the natural stream of daily life provides a foundation for developing theoretical models that explain the underlying processes driving personality expression.

Explaining personality tendencies and dynamics

One can explain personality expression in daily life using structural or dynamic approaches to characterizing individual differences. The structural approach focuses on examining patterns of stability or dispositional tendencies in behaviour and context patterns (Buss & Craik, 1980). In this approach, mobile sensing data can be aggregated from the lower level time units to obtain summary estimates at the between-person level. The dispositional tendencies can then be explained using observational data and/or self-report data representing person factors (e.g. demographics and trait ratings), behavioural factors (e.g. other co-occurring behaviours), and/or contextual factors (e.g. situational cues). Such an approach could also be used to assess the validity of existing personality models (e.g. the Big Five traits and Dark Triad traits) by examining the relationships between personality self-report data and personality sensing data.

The structural approach would serve two primary functions in personality sensing research. First, research examining the construct validity of existing personality models could characterize the strength of the associations between self-report measures and sensed assessments across different levels of aggregation (e.g. at the hourly, daily, weekly, or monthly level). Such analyses would help determine the time points at which the relationship between self-reported traits and sensed tendencies stabilize. Second, the structural approach could yield findings that shed light on the more specific behavioural and situational tendencies that are associated with existing personality models. For example, such analyses of the Big Five trait ratings could determine whether self-reported Extraversion and Conscientiousness (two traits found to be associated with self-reported engagement in physical activity; Rhodes & Smith, 2006) are associated with engagement in naturally observed physical activity tendencies in daily life (e.g. rates of walking, running, and biking). For example, past research in this domain that has focused on dispositional tendencies has shown the Big Five traits map onto mobility behaviours (e.g. distance travelled and places visited; Ai, Liu, & Zhao, 2019), sociability behaviours (e.g. conversation, calling, texting, and social media app use; Harari et al., 2020), and more general mobile phone app use behaviours (e.g. communication, gaming, and transportation; Stachl et al., 2017).

In contrast, the dynamic approach focuses on understanding patterns of change or variability in the behaviour and context patterns. In addition to reflecting mean levels of behaviour, personality is theoretically also expressed in terms of characteristic dynamic patterns of behaviour such as trajectories of behaviour over time or patterns of within-person variability (Fleeson, 2001; Moskowitz & Zuroff, 2004). However, existing methods of personality measurement, which are administered on single or repeated occasions and require participant input, or which refer to general patterns of behaviour, cannot capture such dynamic characteristics over long periods of time and at scale. Personality sensing, with its high-fidelity design, is ideally suited to capturing such dynamic patterns. Sensing data can be analysed at the within-person level to describe within-person variability, examine patterns over time and across situations, and identify within-person structural

dimensions (Harari, Stachl, Muller, & Gosling, in press). For example, past research in this domain has shown that within-person variability in sensed behavioural patterns are associated with the Big Five traits (Wang, Harari, et al., 2018).

Predicting personality and consequential life outcomes

Beyond the description and explanation of personality-relevant information from sensing data, a third empirical domain of personality sensing research is focused on the prediction of personality and other consequential life outcomes. In this domain, machine learning techniques can be used on sensing data to (i) predict a person's personality using scores derived from traditional self-report measures at the trait and/or state level as the ground truth criterion, (ii) predict a person's personality using entirely new dimensions and/or typologies derived from patterns of stability and variability captured in sensor-based observations as the ground truth criterion, and (iii) predict other life outcomes that are theoretically linked to personality using sensing data (e.g. in the domains of well-being and performance; Ozer & Benet-Martínez, 2005).

Personality prediction research presents two core opportunities for informing personality assessment and theory development. First, predictive modelling could be approached in a top-down fashion to help develop automated assessment techniques for the prediction of individual personality scores and life outcomes from sensed data. Second, prediction modelling could be approached in a bottom-up fashion to help map out previously unknown links between sensing data and personality dimensions via the application of interpretable machine learning methods (for further discussion, see Bleidorn & Hopwood, 2019; Stachl et al., 2020). For example, the analyses of prediction models could provide insight into the everyday behavioural and situational patterns that represent manifestations of personality constructs.

It remains to be seen whether personality and life outcome prediction will provide reliable and psychometrically valid assessments that could eventually replace traditional self-report measures. Given that self-reports tap into people's self-views and their subjective perceptions of their behavioural patterns over time, it is quite possible that sensing approaches will instead provide complementary or entirely new sources of personality information. For example, personality prediction research could focus on predicting Extraversion using a variety of potential validation criteria, including self-report scores (e.g. from a Big Five survey), sensed extraverted behaviours (e.g. making calls and communicating via social media and messaging apps), or a combination of the two sources of personality information (e.g. a composite score derived from both self-report and sensing data).

To date, a first wave of personality and life outcome prediction studies, primarily within the field of computer science, have used sensing data to predict self-reported individual personality scores (primarily focusing on the Big Five traits and states; Chittaranjan et al., 2013; Kalimeri, Lepri, & Pianesi, 2013; Mønsted, Mollgaard, & Mathiesen, 2018; De Montjoye, Quoidbach, & Robic, 2013;

Stachl et al., 2019; Teso, Staiano, Lepri, Passerini, & Pianesi, 2013; Wang, Harari, et al., 2018) and life outcomes (primarily focusing on the domains of well-being and performance; Mohr, Zhang, & Schueller, 2017; Wang, Harari, Hao, Zhou, & Campbell, 2015; Wang, Wang, et al., 2018). For the most part, these studies report successful predictions with varying degrees of accuracy for the various traits, states, and life outcomes. However, what remains unclear is the extent to which the findings from past studies are reliable and generalizable because many have been conducted with small to moderate sample sizes (for notable exceptions in the personality prediction domain, see Mønsted et al., 2018; Stachl et al., 2020). Moreover, methodological limitations hinder our current understanding of the extent to which personality and life outcomes can be predicted. For example, the sensed data used to make predictions in any given study often focuses on a narrow range of behaviour (e.g. social interactions or mobility patterns), in part due to the sampling capabilities of the sensing software being used. As a result, the types of sensing data used to make predictions in individual studies often do not overlap across studies, hindering the ability to make cross-study comparisons about the predictors of the outcomes of interest.

In machine learning, the aggregated out-of sample prediction performance of statistical models can be evaluated as an estimator for (i) how well personality-relevant information is encoded in sensing data in general and (ii) how well the prediction models would generalize to other samples to replicate the results. Analytic and methodological issues in the first wave of personality and life outcome prediction research call for further and more extensive prediction attempts to replicate and extend previously reported findings. For example, it seems that early personality prediction studies likely reported inflated predictive performances due to small to moderate sample sizes and performance overestimation due to model overfitting when specifying machine learning models (see Mønsted et al., 2018 for a detailed discussion of this issue). Therefore, additional studies are needed that focus on the prediction of personality scores from larger samples of sensing data and that correctly evaluate the performance of machine learning models (e.g. Mønsted et al., 2018; Stachl et al., 2019). Furthermore, there is a general lack of standardization with regard to best practices when reporting methodological and analytic decisions and a lack of consensus about best approaches for integrating open science practices in sensing research (e.g. providing access to data and/or code needed to reproduce or evaluate findings reported). Many of these current limitations stem from practical considerations surrounding approaches to machine learning for personality prediction and assessment. These practical considerations are covered in depth elsewhere in this special issue, so here, we do not delve into further discussion of these topics. We point interested readers to the Stachl et al. (2020) article in this special issue, which goes into further detail about machine learning techniques for personality prediction.

In sum, all three domains of personality sensing research will provide relevant information about what sensing technologies reveal about personality and how they may be used

to better understand and assess personality expression as it occurs in daily life. Intuitively, we could assume that any digital technology that provides the possibility of recording information about peoples' thoughts and feelings, behaviours, and surrounding context could be used to describe, explain, and predict their personality. However, more systematic and comprehensive studies are needed to build a solid body of knowledge about personality sensing in terms of its potential for furthering scientific research, its viability for use in applied settings, and its broader implications for individuals and society. In the following section, we provide an empirical illustration for two of the three domains of personality sensing research that we have outlined. Specifically, we show how smartphone sensing data can be used to describe and explain personality expression by mapping the relationships between the Big Five traits and everyday behaviours, focusing on physical activity, social behaviour, and mobile phone use.

EMPIRICAL ILLUSTRATION OF PERSONALITY SENSING RESEARCH USING SMARTPHONES

What do smartphones reveal about personality?

What exactly do smartphones reveal about people's personalities? To what extent do they capture patterns of stability and variability in people's everyday thoughts, feelings, behaviours, and situational contexts? What can they teach us about the extent to which people differ from one another and from themselves over time?

Smartphones have become companions to billions of people around the world. According to the Pew Research Center, rates of smartphone ownership have skyrocketed in countries with advanced and emerging economies (Taylor & Silver,2019). For example, in advanced economies, the median percentage of adults who own a smartphone is estimated to be 76%, with some countries nearing full saturation (e.g. 95% of the population in South Korea and 88% in Israel own a smartphone). In emerging economies, the median percentage is 45%, with some countries estimated to have only a quarter of the population owning smartphones (e.g. India's percentage is estimated at 24%). Despite their increased adoption rates, there are disparities in ownership rates such that people who are older, lower in socioeconomic status, and from rural areas are less likely to own a smartphone (Perrin, 2019; Taylor & Silver, 2019). Such adoption rates hold great promise for allowing researchers to move beyond WEIRD (i.e. western, educated, industrialized, rich, and democratic) samples when recruiting participants for their studies (Henrich, Heine, & Norenzayan, 2010), but the current disparities in ownership suggest that biases in sampling may still result in primarily WEIRD samples in the short term. However, as smartphones become more affordable, their adoption rates are likely to further increase in countries around the world.

The pervasive use of smartphones is consequential because smartphones accompany people as they go about their day-to-day lives. The ubiquity of the smartphone seems to be driven by the many and varied functionalities they provide to their owners—they facilitate connection, navigation, information, and entertainment. Moreover, as part of their standard functioning, smartphones can be harnessed to collect immense amounts of personality-relevant information about the individuals who use them.

Smartphones are equipped with a large variety of mobile sensors and metadata logs that continuously sample analogue behavioural and environmental signals in order to enhance the device's functionality. As such, people's everyday behavioural and contextual information are purposely, and in some cases inadvertently, captured in these data streams. For example, text and microphone data collected from smartphones can reveal the language we use in our everyday communication. Accelerometers and phone usage log data can reveal behavioural information about our activities in both physical and digital environments. Ambient light sensors and GPS data can reveal situational information about our surrounding environmental contexts. Moreover, smartphones have several technical features that make them uniquely suited to addressing the limitations of previous approaches to capturing the behaviours and contexts that characterize personality: (i) smartphones already contain many embedded mobile sensors (e.g. accelerometers, microphones, and GPS) and maintain operating system logs that generate metadata (e.g. records of calls, texts, and applications used); (ii) smartphones can be fitted with additional sensing peripherals that collect such parameters as heart rate, temperature, and humidity levels of the surrounding environment; and (iii) smartphones can be used to deliver experience sampling surveys via app notifications, text messages,or emails that probe participants to report on specific information *in situ* (e.g. providing momentary ratings of psychological experiences and situational characteristics).

Despite the ubiquity of smartphones in society and their sophisticated technical features, there has been relatively little empirical research on their potential capabilities for furthering personality theory and assessment. To date, much of the existing research originates from the field of computer science, with few studies conducted by personality psychologists or other social scientists. This scarcity of research is unfortunate because smartphones provide an immense opportunity to add to our understanding of real-world personality expression. The promise of smartphone research for personality psychology is driven by two main opportunities: (i) the potential for data-driven insights to inform existing theories and generate new theoretical approaches and (ii) the ability to scale the assessment of personality by developing unobtrusive and passive measures of personality and related constructs. Moreover, how such assessment technologies come to be used by various stakeholders (e.g. researchers, industry companies, and government organizations) has broad implications for individuals and society. In sum, although many people around the world carry smartphones with them each day, empirical knowledge about the general degree to which personality is revealed in behavioural data collected by smartphones is still limited.

To illustrate the opportunities of smartphone sensing for the understanding of personality, here, we describe and

explain three broad behavioural domains and their relationships with self-reported Big Five personality dimensions. Specifically, we present findings from an exploratory personality sensing study using smartphones to examine (i) descriptive properties that characterize daily behavioural patterns (e.g. average behavioural tendencies and relationships among behavioural patterns at the between-person and within-person levels), (ii) dispositional behavioural tendencies (e.g. stability of behaviour over time, between-person variability, and the extent to which Big Five traits are associated with behavioural tendencies), and (iii) dynamic patterns of change in these behaviours (e.g. within-person variability and relationships between daily behaviour and self-reported Big Five states). Our findings illustrate how each of these approaches can inform our understanding of personality expression in daily life.

Methodological approach

Description of the dataset
Participants were students of a university in the United States who were enrolled in an online introductory psychology course. The data were collected as part of a 2-week self-tracking assignment, in which participants could self-track their psychological experiences (via experience sampling surveys) and behaviours (via smartphone data; for a total of 14 possible days of data collection). Participants could choose to use e-mail (via questions presented using Qualtrics software) or a smartphone sensing application (called CampusLife, which is based on the StudentLife sensing software; Wang et al., 2014) as the self-tracking tool for the class assignment. For the purposes of this article, we focus on the subset of the participants who used the smartphone application, which collected objective data about their behaviours.

Participants also completed a battery of psychological surveys in exchange for personal feedback about their responses. Self-reported personality traits were measured using the 44-item Big Five Inventory (John & Srivastava, 1999. Self-reported personality states were measured during the past hour using a 5-item survey ('During the past hour …'). Specifically, participants were asked: 'How quiet were you?' (extraversion, reversed), 'How considerate and kind were you?' (agreeableness), 'How lazy were you?' (conscientiousness, reversed), 'How anxious and easily upset were you?' (neuroticism), and 'How curious were you?' (openness). The items were rated on a Likert scale using the following 5-point range: 1, not at all; 2, a little bit; 3, somewhat; 4, quite a bit, and 5, very much.

Transparency statements
Our analyses were exploratory and served to illustrate the value of our proposed research agenda. Our sample size was determined based on the number of interested student participants who completed the self-tracking assignment with the smartphone sensing application in the online class. Regarding effect sizes, for the example of a pairwise correlation, our final sample size of 633 participants allowed us to detect effects of $r = .14$, with a statistical power of $1 - \beta = 0.80$, at a significance level of $a = .01$.

The data from this online class have been previously published in other papers that provide additional methodological details about the research design (see Harari, Müller, Mishra, et al., 2017; Harari et al., 2020; Kroencke, Harari, Katana, & Gosling, 2019; Müller et al., in press). The aggregated and de-identified data files and R scripts needed to reproduce our results are available on the OSF at our project page: https://osf.io/xzmkb/.

Assessing everyday behaviours with a smartphone sensing application
The CampusLife application was designed to run on both Android and iOS phones. Below, we describe how the behaviours were assessed using both on-device classification of behaviour and off-device data processing of the mobile sensors and metadata logs.

Inferring physical activity from accelerometers. Physical activity was measured using activity recognition classifiers developed by Google for Android phones (Google Activity Recognition API, 2017) and by Apple for iOS phones (Apple Core Motion, 2017), which use accelerometer data to detect activity-based behavioural inferences (e.g. stationary, walking, on a bicycle). Using these activity classifiers we were able to infer users' duration of time spent stationary, moving, walking, running, and biking. We used these activity inferences to aggregate the data into duration of time spent physically active for each hour of each day in the data collection period. This behavioural estimate captured several aspects of physical activity behaviour—the tendency to move (vs. be stationary) and more narrow types of physical activity behaviour as indexed by the amount of time participants spent walking, running, or biking. It should be noted that these physical activity estimates may be underestimating movement behaviours and/or overestimating stationary behaviours because the classifier is capturing the duration of time in which the phone itself is stationary or moving. So, if a participant were to leave their phone behind while walking around or exercising, we would not be able to capture such information accurately (resulting in an overestimate of their stationary behaviour and an underestimate of their walking behaviour).

Inferring conversation behaviours from microphone sensors. As described in Harari et al. (2020), conversation was measured using an audio classifier developed in prior work (Lane et al., 2012; Rabbi et al., 2011); the classifier achieved 84–94% accuracy at classifying microphone data into audio-based inferences (i.e. silence, noise, and voices). Every third minute, the microphone was sampled and a classifier inferred whether voices were detected. The voice inferences were used to derive estimates of the duration of time spent around other voices (vs. silence or noise) and the frequency of separate instances of conversation (Wang et al., 2014). If a conversation (i.e. voices) was detected, the microphone continued to be sampled until the conversation was over. The application saved the audio inferences as a '0' for silence, '1' for noise, '2' for voices, and '3' for unknown, so the content of conversations was never recorded. The audio inferences were compiled to create variables for the

duration of time spent proximal to human speech (either in conversation or around conversation) for each hour of each day monitored. This behavioural estimate captured a unique aspect of social behaviour—the tendency to affiliate with others as indexed by the amount of time participants spent around conversations and around separate instances of conversation (capturing both when a participant was actively engaged in a conversation and around conversations).

Inferring mobile phone use from metadata logs

Mobile phone use was measured using metadata logs that recorded time-stamped events denoting each time the phone was locked or unlocked. We used the unlock event logs to infer the amount of mobile phone use by aggregating the data into frequency and duration of time spent using the phone for each hour of each day in the data collection period. This behavioural estimate captured a unique aspect of mobile phone use behaviour—the general tendency to use one's mobile phone as indexed by the number and amount of time the phone's screen was in an unlocked state.

Data processing

To obtain the daily-level behavioural estimates, we estimated the amount of behaviour engaged in (frequency and duration) by summing up the observations within each day based on timestamps associated with the behavioural records collected by the sensing app. Our data processing steps followed this general order for each behaviour, per person: (1) We wanted to ensure the estimates were representative estimates of the participants' behaviour for each day, so we created a threshold for the minimum number of hours of sensor data needed per day (>14 h or over 60% of the day) for the data to be retained in the analyses. This threshold was used in the data-cleaning process to identify and remove any days with an insufficient amount of hourly data per participant; (2) The daily estimates were then computed on the retained data by summing across the 24 h within each day to obtain a behavioural estimate per day for each participant; and (3) We retained participants who had more than 1 day of data across the sensing data modalities.[1]

Finally, to handle outliers and address linear modelling assumptions, we also applied winsorization (Ghosh & Vogt, 2012) to the daily duration estimates to replace extreme values. The winsorization of the duration variables helped improve the modelling assumptions and the Akaike information criterion values (compared with standard models run with the extreme outliers included), so we retained the data and models with the winsorized duration estimates.

To obtain estimates of people's dispositional tendencies, we performed an additional aggregation step Specifically, we aggregated the data to compute a within-person average estimate that represents an individual's behavioural tendency across days. This process produced for each participant, a single behavioural tendency estimate for each of the specific behaviours: stationary duration, movement duration, walking duration, running duration, biking duration, conversation frequency, conversation duration, phone unlock frequency, and phone unlock duration. These tendency estimates were used in a series of regression analyses reported in the section on examining dispositional behavioural tendencies.

EXPLORATORY RESULTS MAPPING SENSED BEHAVIOURS TO THE BIG FIVE

Describing the everyday behaviours of young adults

To provide a descriptive understanding of everyday behaviour patterns, we started by examining average behavioural tendencies to describe the behaviour of a typical person on a typical day. Second, we examined between and within person correlations to describe inter and intra individual differences in daily behaviours.

Behaviour of a typical person on a typical day (average behavioural tendencies)

To describe the daily behavioural patterns of young adults, we first computed the within-person averages across days to obtain a daily behavioural tendency score for each individual. We then computed the between-person means of these behavioural tendencies to examine the average behavioural tendencies for a typical person on a typical day. Our descriptive findings chart the everyday behaviours of young adults across the three behavioural domains (see Table 1). The physical activity tendencies revealed that on average, the typical young adult spent approximately 18.4 h stationary, 2.8 h moving, 1.4 h walking, and only 0.69 min running and 9.5 min biking per day. The conversation behaviour tendencies revealed that on average, young adults had around 20 conversations for an average daily duration of 2.5 h. The mobile phone usage tendencies revealed that on average, the typical young adult unlocked their phone 72 times and spent 2.4 h on their phone per day.

Inter-individual differences in daily behaviours (between-person correlations)

To examine the relationships between people's daily behaviour patterns, we computed between-person Spearman's correlations among the daily behavioural estimates (see half of Table 2 below the diagonal). For physical activity, the findings show (as one would expect) that people who tend to move more in general, also tend to walk, run, and bike more (r's range from .20 to .66) and spend less time stationary ($r = -.44$), compared with people who tend to move less. People who tend to move more also tend to be around more frequent and longer conversations (r's are .27 and .24, respectively) and use their phone more frequently ($r = .24$), compared with people who tend to move less.

The between-person correlations between the social behaviours and mobile phone use behaviours show that people who tend to be around more frequent and longer conversations also tend to use their phone more frequently (r's are .44 and .37, respectively) but for lower durations of time ($r = -.17$ and $-.14$, respectively).

---

[1]This process left us with a final sample of 633 participants who had more than 1 day of sensing data.

Table 1. Descriptive information for sensed daily behaviours of young adults

	M_{AVG}	SD_{AVG}	Min	Med	Max	Skew	Kurtosis
Physical activity behaviours							
Stationary duration	1104.81	134.85	638.85	1116.20	1392.42	−0.67	0.35
Movement duration	167.27	73.23	1.98	158.29	443.97	0.54	0.27
Walking duration	81.89	43.10	0.00	77.50	246.02	0.44	0.42
Running duration	0.69	1.25	0.00	0.23	16.68	5.95	60.93
Biking duration	9.55	14.13	0.00	2.23	119.73	2.21	7.66
Social behaviours							
Conversation frequency	19.92	13.86	0.00	18.39	209.00	4.86	61.04
Conversation duration	148.37	104.18	0.00	132.69	586.23	0.70	−0.01
Phone use behaviours							
Unlock frequency	72.25	54.07	0.00	57.88	344.00	1.64	3.66
Unlock duration	143.80	97.64	0.00	129.82	480.00	0.76	0.21
Personality states							
Extraversion	2.72	0.56	1.00	2.75	4.30	−0.23	0.04
Agreeableness	3.53	0.80	1.21	3.52	5.00	−0.11	−0.47
Conscientiousness	3.51	0.65	1.00	3.55	4.92	−0.37	−0.04
Neuroticism	1.91	0.70	1.00	1.78	5.00	1.10	1.55
Openness	2.47	0.78	1.00	2.44	5.00	0.37	−0.14

Note. $N = 633$ for physical activity, social, and phone use behaviours, and N for personality states ranges from 624 to 629. Duration estimates are in minutes. Frequency estimates are in counts. Personality states represent daily states.

Intra-individual differences in daily behaviours (within-person correlations)

To examine the relationships within people's daily behaviour patterns, we computed within-person Spearman's correlations among the daily behavioural estimates (see the half of Table 2 above the diagonal). For physical activity behaviours, the findings show that on days when people had moved more in general, they also tended to show more walking, running, biking (r's range from .16 to .71), and less stationary behaviour ($r = −.51$). On days when people spent more time moving, they also tended to have more frequent and longer conversations (r's are .33 and .35, respectively), compared with days when they moved less. Moreover, on days when people moved more in general, they also used their phone more frequently ($r = .23$), but for lower durations ($r = −.05$), compared with days when they moved less.

The within-person correlations between the social behaviours and mobile phone use behaviours were similar in pattern to the between-person correlations. Specifically, the findings show that on days when people had more frequent and longer conversations, they also tended to use their phone more frequently (r's are .29 and .23, respectively), but for lower durations of time (r's are −.10 and −.08, respectively), compared with days when fewer and lower durations of conversation were detected.

Examining dispositional behavioural tendencies

To explain everyday behaviour patterns, we started by examining between-person variability to identify the amount of variance due to individuals (i.e. person factors). Second, we examined mean-level consistency in the daily behaviours to determine the degree of stability over time. Third, we

conducted regression analyses to explain the observed behavioural tendencies using self-reported Big Five trait scores.

Amount of variance due to individuals (between-person variability)

To determine how much of the total variance in the observed behavioural patterns was due to between-person factors, we computed intra-class correlation coefficient (ICC) (1,1) estimates for each of the behaviours across the three domains. As shown in Table 3, the highest ICC(1) estimates based on their confidence intervals were observed for the number of phone unlocks (.64) and the amount of conversation frequency (.60) across days, suggesting that a great deal of the variance observed in this sample for these two behaviours can be attributed to individual differences. The lowest estimates were observed for the amount of time spent running (.16) and using the phone (.19), suggesting that variability observed in these behavioural estimates may be driven by other factors (e.g. within-person factors, contexts, time, and error).

Degree of stability across days (mean-level consistency)

To examine stability in the day-to-day behavioural patterns, we computed the ICC(3,*k*) estimates to understand the extent to which the behaviours were consistent over time. As shown in Table 3, the consistency estimates for the behaviours across days were generally high. Based on their confidence intervals, the highest estimates were observed for the number of times the phone was unlocked (.96) across days, the number and duration of conversations (.95 and.94), and the amount of time spent biking (.90), walking (.88), and moving (.87). These high consistency estimates show that young adults are quite stable in their mean level of engagement in these behaviours over time. In contrast, the lowest estimates

Table 2. Between-person and within-person correlations among the sensed daily behaviours

	1	2	3	4	5	6	7	8	9
1. Stationary duration		**−.51** [−.53, −.48] < .001	**−.31** [−.34, −.28] < .001	−.03 [−.07, .01] .154	**−.15** [−.18, −.11] < .001	**−.04** [−.07, .00] .030	**−.09** [−.12, −.05] < .001	**.05** [.02, .08] .004	**−.15** [−.18, −.12] < .001
2. Movement duration	**−.44** [−.51, −.37] < .001		**.71** [.69, .73] < .001	**.16** [.12, .20] < .001	**.32** [.29, .35] < .001	**.33** [.3, .36] < .001	**.35** [.32, .38] < .001	**.23** [.20, .26] < .001	**−.05** [−.08, −.01] .010
3. Walking duration	**−.38** [−.45, −.31] < .001	**.66** [.61, .71] < .001		**.22** [.18, .26] < .001	**.25** [.22, .28] < .001	**.28** [.25, .31] < .001	**.26** [.23, .29] < .001	**.26** [.23, .29] < .001	**−.07** [−.10, −.03] < .001
4. Running duration	−.04 [−.14, .05] .381	**.20** [.11, .29] < .001	.47 [.4, .54] < .001		**.12** [.08, .16] < .001	**.10** [.06, .14] < .001	**.11** [.07, .15] < .001	**.13** [.09, .17] < .001	−.01 [−.05, .03] .689
5. Biking duration	**−.17** [−.25, −.09] < .001	**.45** [.38, .52] < .001	.31 [.24, .39] < .001	.05 [−.05, .14] .277		**.16** [.13, .20] < .001	**.17** [.14, .20] < .001	**.14** [.10, .17] < .001	.00 [−.04, .03] .826
6. Conversation frequency	.05 [−.04, .13] .278	**.27** [.19, .35] < .001	.15 [.07, .23] < .001	**.28** [.19, .36] < .001	**.15** [.07, .23] < .001		**.64** [.62, .66] < .001	**.29** [.26, .32] < .001	**−.10** [−.14, −.07] < .001
7. Conversation duration	.03 [−.05, .12] .412	**.24** [.16, .32] < .001	.13 [.05, .21] .002	**.27** [.18, .36] < .001	**.12** [.03, .20] .005	**.92** [.91, .94] < .001		**.23** [.20, .27] < .001	**−.08** [−.12, −.05] < .001
8. Unlock frequency	**.18** [.10, .26] < .001	**.24** [.16, .32] < .001	.03 [−.05, .11] .464	**.22** [.13, .3] < .001	**.13** [.05, .21] .001	**.44** [.37, .5] < .001	**.37** [.30, .44] < .001		**.11** [.08, .13] < .001
9. Unlock duration	**−.32** [−.40, −.25] < .001	−.05 [−.13, .04] .247	.06 [−.02, .14] .123	.04 [−.05, .13] .288	.01 [−.07, .09] .835	**−.17** [−.25, −.09] < .001	**−.14** [−.22, −.06] < .001	**−.20** [−.27, −.12] < .001	

Note. N = 633 participants, and N = 5,414 observations. Spearman correlations and corresponding 95% confidence intervals and p values. Within-person correlations are depicted above the diagonal. Between-person correlations are depicted below the diagonal. Coefficients in bold are significant with p ≤ .05.

Table 3. Variability and consistency estimates for sensed daily behaviours of young adults

	Between-person variance ICC(1)	Single measurement consistency ICC(3,1)	Multiple measurement consistency ICC(3,k)
Physical activity behaviours			
Stationary duration	.28 [.26, .31]	.29 [.27, .32]	.83 [.82, .85]
Movement duration	.26 [.24, .29]	.28 [.26, .31]	.83 [.81, .84]
Walking duration	.37 [.34, .39]	.38 [.35, .41]	.88 [.87, .89]
Running duration	.16 [.14, .17]	.16 [.14, .18]	.69 [.66, .72]
Biking duration	.43 [.40, .46]	.43 [.41, .46]	.90 [.89, .91]
Social behaviours			
Conversation frequency	.60 [.57, .62]	.61 [.58, .63]	.95 [.94, .95]
Conversation duration	.53 [.51, .56]	.55 [.52, .57]	.94 [.93, .94]
Phone use behaviours			
Unlock frequency	.64 [.62, .66]	.65 [.63, .67]	.96 [.95, .96]
Unlock duration	.19 [.17, .21]	.19 [.17, .21]	.74 [.71, .76]
Personality states			
Extraversion	.22 [.20, .24]	.23 [.21, .25]	.82 [.80, .83]
Agreeableness	.58 [.55, .60]	.60 [.58, .62]	.96 [.95, .96]
Conscientiousness	.34 [.31, .36]	.35 [.33, .37]	.89 [.88, .90]
Neuroticism	.48 [.46, .51]	.49 [.46, .51]	.93 [.93, .94]
Openness	.53 [.51, .56]	.55 [.52, .57]	.95 [.94, .95]

Note. The variability and consistency estimates were computed using the 'psych' package in R and are based on $N = 633$ participants with $N = 5414$ observations. The between-person variance represents the ICC(1,1), which is interpreted as the per cent of variation in the observed daily behavioural estimates that can be explained by individual factors. The single measurement consistency estimates represent the ICC(3,1), with the following specification: two-way mixed effects, consistency, and single measurement (i.e. day). The mean measurement consistency represents the ICC(3, k), with the following specification: two-way mixed effects, consistency, and multiple measurements (where $k = 12$ days). The mean measurement consistency is interpreted as the average stability of the daily behavioural assessments across days.

were observed for the amount of time spent running (.69), suggesting that young adults engage in this behaviour at relatively less stable rates from day to day. Such findings permit comparisons among behaviours with regard to which may be more or less stable over time and suggest that mean levels of physical activity, conversation, and phone use behaviours generally show consistency from day to day.

Explaining behavioural tendencies from traits (regression models)

Next, we report on a series of regression analyses that examine the explanatory ability of the established Big Five trait scores with respect to the sensed behavioural tendency measures. Specifically, we explain people's daily behavioural tendencies (i.e. their within-person means aggregated across the 2-week period) from their self-reported Big Five trait scores. Given that our data consisted of both duration and frequency data, we specified a series of linear regressions (for the duration variables) and negative binomial regression models (for the frequency variables) when using the Big Five trait ratings to explain the behavioural tendencies (see Table 4 for all nine models).

Controlling for the effects of age and sex, the Big Five traits were significantly associated with six of the nine behavioural tendencies we studied here, explaining 1% to 7% of the variance in the daily behavioural tendencies. The physical activity and conversation behavioural tendencies were best explained by Big Five traits, such as the amount of time

a person tends to spend walking per day ($R^2 = .07$) and the amount of time a person tends to spend in conversation per day ($R^2 = .04$). In contrast, the behavioural tendencies that were not associated with the Big Five personality traits were the daily running and biking[2] and the frequency of phone unlocking tendencies. Below, we describe our findings across all nine models for each trait in turn.

Extraversion was associated with behavioural tendencies in the domains of physical activity behaviour and social behaviour. Specifically, extraversion was positively associated with movement tendencies, and walking tendencies, as well as negatively related to being stationary. Extraversion was also positively associated with conversation frequency and duration tendencies. These findings suggest that people who reported being more extraverted also tended to generally engage in greater amounts of physical activity behaviour, and more frequent and longer conversation behaviour, compared with people who reported being lower on extraversion.

Agreeableness was not associated with the behavioural tendencies in the domains of physical activity behaviour, social behaviour, or mobile phone use behaviour.

[2]Despite the various data processing and transformation steps we pursued, the running and biking behavioural tendencies remained non-normal due to low base rates for these behaviours (i.e. positively skewed). As such, the model results for these two variables should be interpreted with caution because the models violated several assumptions (see our online materials for more details).

Table 4. Explaining sensed behavioural tendencies from sex, age, and Big Five personality traits

| | Physical activity behaviours | | | | | | | | |
| | Stationary duration (df = 436) | | | Movement duration (df = 436) | | | Walking duration (df = 458) | | |
	Est.	95% CI	p	Est.	95% CI	p	Est.	95% CI	p
Intercept	1042.23	[997.94, 1086.52]	<.001	210.90	[186.09, 235.71]	<.001	114.09	[99.89, 128.29]	<.001
Sex	**42.75**	[16.73, 68.77]	.001	**−28.57**	[−43.15, −14.00]	<.001	**−20.14**	[−28.48, −11.80]	<.001
Age	8.70	[−2.13, 19.54]	.115	0.55	[−5.52, 6.62]	.858	**−3.95**	[−7.46, −0.44]	.028
Extraversion	**−23.36**	[−36, −10.71]	<.001	**12.93**	[5.84, 20.01]	<.001	**8.26**	[4.25, 12.28]	<.001
Agreeableness	−8.97	[−21.54, 3.60]	.161	−2.79	[−9.83, 4.25]	.436	−1.67	[−5.71, 2.37]	.418
Conscientiousness	5.88	[−6.51, 18.27]	.351	2.73	[−4.21, 9.67]	.441	**4.11**	[0.19, 8.04]	.040
Neuroticism	0.00	[−13.54, 13.54]	1.00	0.90	[−6.68, 8.49]	.815	3.16	[−1.16, 7.49]	.151
Openness	−9.97	[−21.98, 2.04]	.103	0.24	[−6.49, 6.97]	.944	−0.57	[−4.42, 3.28]	.771
Adjusted R^2	.06			.05			.07		

| | Running duration (df = 363) | | | Biking duration (df = 458) | | |
	Est.	95% CI	p	Est.	95% CI	p
Intercept	1.58	[1.07, 2.10]	<.001	11.23	[6.48, 15.97]	<.001
Sex	**−0.53**	[−0.83, −0.23]	.001	−1.37	[−4.15, 1.42]	.335
Age	−0.09	[−0.21, 0.03]	.135	0.26	[−0.91, 1.44]	.658
Extraversion	0.13	[−0.01, 0.27]	.077	1.01	[−0.33, 2.35]	.138
Agreeableness	0.02	[−0.12, 0.16]	.779	−0.51	[−1.86, 0.84]	.456
Conscientiousness	0.01	[−0.13, 0.15]	.878	0.91	[−0.40, 2.22]	.171
Neuroticism	0.02	[−0.13, 0.17]	.818	0.01	[−1.44, 1.45]	.994
Openness	0.07	[−0.07, 0.20]	.317	−0.12	[−1.41, 1.16]	.852
Adjusted R^2	.03			.00		

| | Social behaviours | | | | | |
| | Conversation frequency (df = 455) | | | Conversation duration (df = 455) | | |
	Est.	95% CI	p	Est.	95% CI	p
Intercept	3.10	[2.87, 3.34]	<.001	178.28	[143.52, 213.04]	<.001
Sex	−0.08	[−0.22, 0.06]	.267	−19.37	[−39.80, 1.07]	.063
Age	−0.04	[−0.09, 0.03]	.232	−3.47	[−12.06, 5.12]	.427
Extraversion	**0.15**	[0.08, 0.22]	<.001	**22.77**	[12.82, 32.72]	<.001
Agreeableness	−0.02	[−0.09, 0.04]	.476	−5.75	[−15.72, 4.22]	.258
Conscientiousness	0.03	[−0.04, 0.10]	.345	5.30	[−4.38, 14.98]	.283
Neuroticism	**0.08**	[0.01, 0.15]	.031	9.14	[−1.55, 19.82]	.094
Openness	0.01	[−0.05, 0.08]	.734	−0.50	[−10.06, 9.07]	.919
2 × log likelihood	−3590.33					
Pseudo R^2/Adjusted R^2	.05			.04		

| | Mobile phone behaviours | | | | | |
| | Unlock frequency (df = 508) | | | Unlock duration (df = 508) | | |
	Est.	95% CI	p	Est.	95% CI	p
Intercept	4.52	[4.28, 4.76]	<.001	125.18	[92.70, 157.66]	<.001
Sex	−0.14	[−0.28, 0.00]	.054	11.28	[−7.68, 30.23]	.243
Age	**−0.08**	[−0.14, −0.02]	.009	−5.77	[−14.00, 2.45]	.168
Extraversion	0.05	[−0.01, 0.12]	.117	3.89	[−5.26, 13.05]	.404
Agreeableness	−0.01	[−0.08, 0.05]	.724	6.78	[−2.34, 15.90]	.145
Conscientiousness	−0.04	[−0.11, 0.03]	.240	**−11.40**	[−20.34, −2.47]	.013
Neuroticism	0.00	[−0.07, 0.08]	.950	−4.57	[−14.29, 5.15]	.356
Openness	−0.04	[−0.10, 0.03]	.268	0.63	[−8.12, 9.39]	.887
2 × log likelihood	−5373.57					
Pseudo R^2/Adjusted R^2	.03			.01		

Note. Est. = estimates from linear (duration variables) and negative binomial regression (frequency variables) models. DVs = daily tendencies of engaging in different behaviours (daily estimates averaged across days). IVs = personality traits (z standardized). All models controlled for sex (0 = male, 1 = female) and age (z standardized). We report the Adjusted R^2 for the linear models (duration variables) and 2× log likelihood as well as pseudo R^2 (Cox and Snell Index) for the negative binomial regressions (frequency variables). Coefficients in bold are significant with $p \leq .05$.

Conscientiousness was associated with behavioural tendencies in the domain of physical activity behaviour and mobile phone use behaviour. Specifically, people who reported higher conscientiousness tended to generally walk more and engage in less time using their mobile phone compared with people who reported being lower on conscientiousness.

Neuroticism was associated with behavioural tendencies in the domain of social behaviour. Specifically, people who reported higher neuroticism tended to have more frequent conversations compared with people who reported being lower on neuroticism.

Openness to experience was not associated with the behavioural tendencies in the domains of physical activity behaviour, social behaviour, or mobile phone use behaviour.

Examining dynamic patterns of behavioural change

To examine dynamic patterns in everyday behaviour, we started by visualizing density distributions to characterize variability in the behaviours over time. Second, we conducted multilevel models to explain changes in self-reported personality state scores from sensed daily behaviours.

Characterizing variability in behaviour over time (density distributions)

To highlight the variability in behaviour over time, we first present density distributions for a few participants in our sample that portray the variability both between and within persons. Figure 3 presents the density distributions for three participants who were randomly selected among those who had more than 7 days of data for all behaviours. This
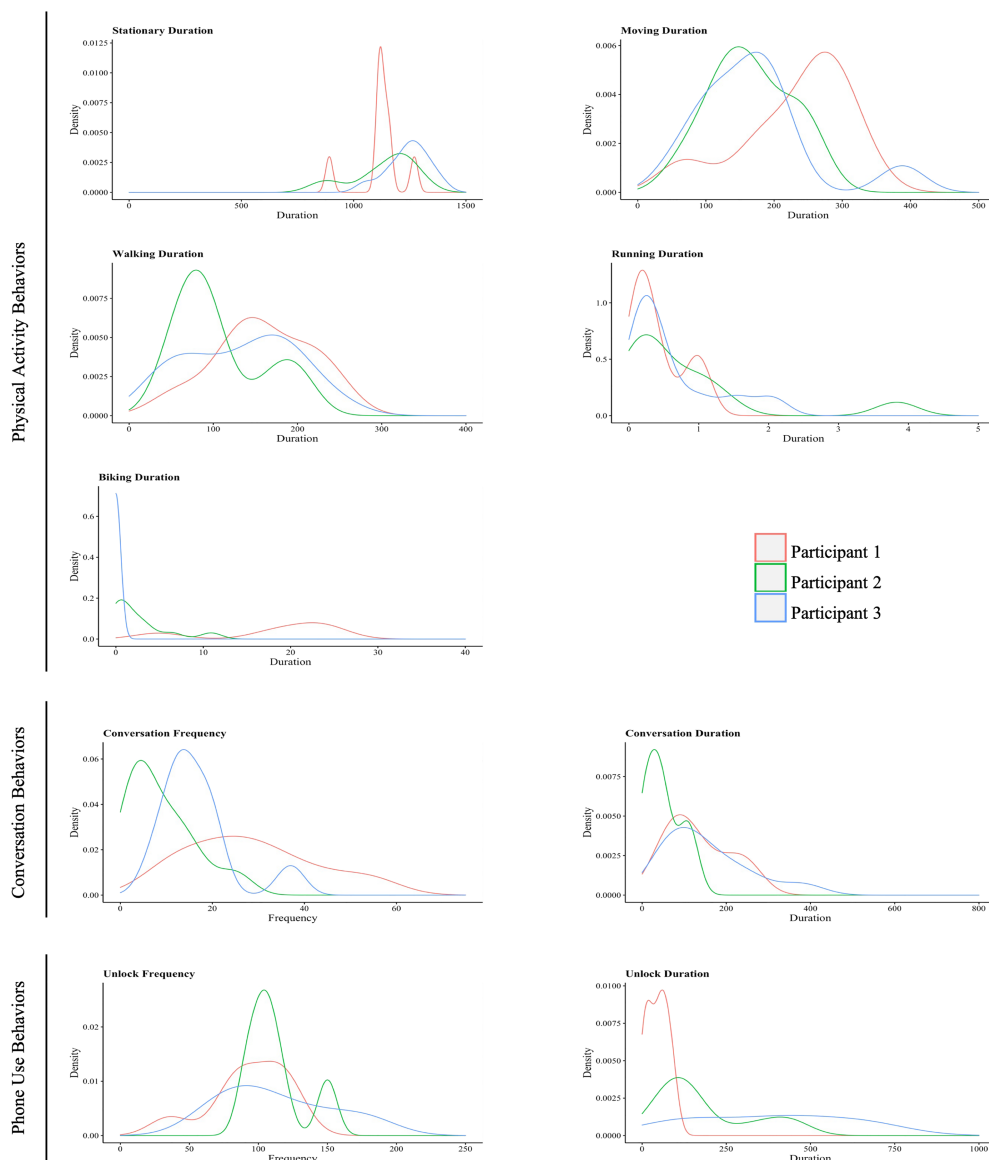

FIGURE 3.   Density distributions for the daily behaviours of three participants for physical activity, conversation, and phone use. Duration estimates are in minutes. Frequency estimates are in counts. Density values refer to the probability density function and add up to 1.

approach allows us to visualize the extent to which individuals vary from one another in their behavioural patterns over time (by comparing the shapes of the individual distributions) and vary from themselves over time (by examining the width of the distributions for any given individual). Moreover, the within-person summary statistics provide quantitative information about the shape of a person's behavioural distribution over time.

Explaining self-reported Big Five states from daily behaviour (multilevel models)
To understand the relationship between people's everyday behaviour and their personality states, we report on a series of multilevel models that examine the extent to which the behaviours people engage in are associated with their daily Big Five states. Specifically, we modelled the relationships between people's average daily Big Five state scores (as

dependent variables) and their daily sensed behaviours (as independent variables) in a series of random-intercept fixed slope multilevel models in which days were nested within person (see Table 5 for all five models).

Daily extraversion states were associated with the amount of time spent stationary and walking, the frequency and duration of conversations, and the duration of mobile phone use per day. These findings suggest that compared with their average daily behavioural tendencies, people reported to be more extraverted on days when they spent less time stationary, more time walking, were around more and longer conversations, and spent less time using their mobile phone.

Daily agreeableness states were not associated with the physical activity, social, or mobile phone use behaviours engaged in per day. These findings suggest that daily-level change in considerate and kind states were not associated with these behaviours at the daily level.

Table 5. Explaining daily personality states from sensed daily behaviours

	Extraversion (N = 368, Obs = 1862)			Agreeableness (N = 362, Obs = 1,793)			Conscientiousness (N = 368, Obs = 1861)		
	Est.	95% CI	p	Est.	95% CI	p	Est.	95% CI	p
Intercept	2.81	[2.75, 2.88]	<.001	3.49	[3.40, 3.58]	<.001	3.56	[3.49, 3.64]	<.001
Physical activity behaviours									
Stationary duration	**−0.05**	[−0.10, 0.00]	.031	−0.01	[−0.05, 0.03]	0.748	**−0.08**	[−0.13, −0.03]	.001
Movement duration	**0.10**	[0.03, 0.16]	.004	0.03	[−0.02, 0.08]	0.32	−0.05	[−0.11, 0.01]	.109
Walking duration	0.03	[−0.02, 0.09]	.238	0.02	[−0.03, 0.06]	0.423	**0.13**	[0.07, 0.18]	<.001
Running duration	0.00	[−0.04, 0.04]	.986	−0.03	[−0.06, 0.01]	0.107	0.03	[−0.01, 0.07]	.111
Biking duration	0.05	[0.00, 0.09]	.064	0.02	[−0.01, 0.06]	0.219	−0.02	[−0.07, 0.02]	.329
Social behaviours									
Conversation frequency	**0.10**	[0.05, 0.16]	<.001	0.03	[−0.01, 0.07]	0.199	0.03	[−0.02, 0.08]	.225
Conversation duration	**0.13**	[0.07, 0.18]	<.001	0.00	[−0.04, 0.04]	0.998	0.01	[−0.05, 0.06]	.827
Phone use behaviours									
Unlock frequency	**−0.07**	[−0.12, −0.02]	.009	0.00	[−0.05, 0.04]	0.877	0.04	[−0.01, 0.09]	.126
Unlock duration	**−0.06**	[−0.10, −0.01]	.009	−0.03	[−0.07, 0.00]	0.052	**−0.05**	[−0.09, −0.01]	.015

	Neuroticism (N = 367, Obs = 1850)			Openness (N = 367, Obs = 1847)		
	Est.	95% CI	p	Est.	95% CI	p
Intercept	1.88	[1.80, 1.96]	<.001	2.44	[2.35, 2.53]	<.001
Physical activity behaviours						
Stationary duration	−0.02	[−0.06, 0.01]	0.233	−0.01	[−0.05, 0.03]	0.639
Movement duration	−0.03	[−0.08, 0.02]	0.247	−0.05	[−0.11, 0.00]	0.069
Walking duration	0.00	[−0.04, 0.05]	0.838	**0.07**	[0.02, 0.12]	0.005
Running duration	0.00	[−0.03, 0.04]	0.798	0.02	[−0.01, 0.06]	0.242
Biking duration	0.03	[0.00, 0.02]	0.293			
Social behaviours						
Conversation frequency	−0.04	[−0.08, 0.00]	0.054	0.03	[−0.02, 0.07]	0.261
Conversation duration	−0.02	[−0.06, 0.02]	0.298	−0.04	[−0.09, 0.00]	0.073
Phone use behaviours						
Unlock frequency	0.02	[−0.02, 0.06]	0.354	0.00	[−0.05, 0.04]	0.914
Unlock duration	−0.01	[−0.05, 0.02]	0.393	−0.01	[−0.04, 0.03]	0.683

Note. Est = estimates from multilevel regression models. DVs = aggregated daily personality state scores. IVs = within-person scaled daily behavioural estimates. Coefficients in bold are significant at $p \leq .05$.

Daily conscientiousness states were associated with the amount of time spent stationary and walking and the duration of mobile phone use per day. These findings suggest that compared with their average daily behavioural tendencies, people reported to be more conscientious on days when they spent less time stationary, more time walking, and spent less time using the phone.

Daily neuroticism was not associated with the physical activity, social, or mobile phone use behaviours engaged in per day. These findings suggest that daily-level change in anxiety states was not associated with these behaviours at the daily level.

Daily openness states were associated with the amount of time spent moving and walking. These findings suggest that compared with their average daily behavioural tendencies, people reported to be more open on days when they spent less time moving in general, but more time walking.

Summary of results

Taken together, the findings from our empirical illustration show the value of personality sensing research for furthering our understanding of everyday behaviours and their relationship to personality expression at the trait and state levels. Specifically, the descriptive findings provide an illustrative portrait of the behavioural lifestyles of young adults, identifying relationships among the behaviours between and within persons. The analyses of dispositional behavioural tendencies show that self-reported Big Five states were associated with daily tendencies to engage in physical activity, social behaviour, and mobile phone use (R^2 ranged from .01 to .07 across the models for the duration estimates, and the pseudo R^2 values were .02 and .05 for the frequency estimates). The analyses of dynamic patterns show that people's behavioural patterns showed a great deal of between-person and within-person variability.

Moreover, the findings from our multilevel models show that at the daily level, behaviours also explain a small proportion of the variability in daily personality state expression. Overall, the findings are noteworthy in that they show that (i) sensing technologies provide fine-grained estimates of real-world behaviour that can be used to describe people's behavioural patterns at the between and within person levels; (ii) self-reported traits are associated with observed behavioural tendencies, lending support for the construct validity of the Big Five; and (iii) day-to-day change in behaviours are associated with self-reported changes in perceived personality state expression over time.

OUTLOOK

There are many opportunities for personality sensing research to transform the understanding and assessment of personality in psychological science. Having outlined three of the core empirical domains of opportunity—for description, explanation, and prediction of personality expression—and empirically illustrated how such an agenda could be applied in a smartphone-based personality sensing study, we conclude with a brief discussion of considerations for personality sensing research moving forward. We focus our outlook on considerations for (i) understanding and assessing personality expression in daily life, (ii) establishing methodological standards and best practices for personality sensing research, and (iii) using sensing technologies to move towards a behavioural personality science.

Understanding and assessing personality expression in daily life

The ability to simultaneously capture people's behaviour, thoughts and feelings, and situations in personality sensing research points to several directions for understanding and assessing personality. First, a better theoretical understanding of personality can be obtained by examining everyday behavioural acts and dispositional tendencies (e.g. Buss & Craik, 1980); momentary personality state expression (e.g. Fleeson, 2001); and the relationships between people, their behaviour, and the environment (e.g. Funder, 2006; Russell & Ward, 1982). Moreover, personality sensing is poised to expand our understanding of personality expression in different cultures by permitting easier recruitment of individuals from non-WEIRD societies into research studies (Henrich, Heine, & Norenzayan, 2010). People around the world use sensing technologies, and the adoption rates will continue to increase as the technology becomes more affordable. The widespread diffusion of sensing technologies permits researchers to recruit participants from around the world by having them download sensing software to their own personal devices, which can then be used to conduct a broad array of theoretically informed psychological studies.

Second, personality sensing research promises to enable the development and evaluation of passive personality assessment. The first wave of research focused on personality prediction from sensing data has demonstrated the viability of passive assessment. However, more research is needed before personality assessment models derived from sensing data should be used in applied settings. To evaluate newly developed personality assessments (e.g. prediction and classification models), the field performance of behavioural, contextual, and psychological models in ecologically valid, real-world settings will be needed. After the models are developed, research is needed to evaluate them in intensive test-bed studies that assess the performance, generalizability, and validity of the developed personality assessment models. For example, we suggest evaluating the generalizability of the personality classifiers and examining the level of accuracy in prediction models over time to determine at what point the personality estimates stabilize (e.g. 1 day, 3 days, 1 week, or 1 month) and across countries to determine how prediction models perform in different cultures (e.g. Khwaja et al., 2019). Such an approach to validating the personality assessment models could also be used to validate any pre-existing and newly developed behavioural or contextual classifiers that are being used for assessment of personality-relevant information.

Establishing best practices for methodological and ethical standards

While deriving personality-relevant information from sensing data may seem like a straightforward approach, the advent of widespread sensing technologies has only recently enabled this new frontier in personality science. Whereas traditional approaches to studying personality (e.g. self-reports) have established guidelines for data collection and a set of best practices for creating variables and modelling personality data, many potential researcher degrees of freedom are introduced when trying to empirically study personality-relevant constructs using sensing technologies. There are numerous unresolved methodological considerations and ongoing discussions about creating a set of guidelines for conducting interdisciplinary sensing research. Such guidelines will benefit most from establishing interdisciplinary collaborations, intellectual communities (e.g. the LifeSensing Consortium and MD2K[3]), and opportunities for cross-disciplinary training that go beyond the scope of the training that can be provided in psychological science alone. Empirical research on personality sensing cannot move forward systematically without a discussion of methodological standards and best practices to guide the field.

Mobile sensing as a research method will undoubtedly help to overcome many current challenges in personality science. However, ongoing challenges such as the use of unrepresentative samples (Henrich, Heine, & Norenzayan, 2010) or the replicability of research results (Open Science Collaboration, 2015) will not be resolved by the use of sensing methods. Rather, sensing methods merely offer improved conditions (e.g. opportunities for online recruitment, collecting objective behavioural data, high ecological validity, and granular observations) to facilitate the implementation of higher scientific standards for personality research. Consequently, researchers will need to (i) adopt an explicit, empirical set of standards for acceptable mobile sensing data quality and reporting standards for outlining processing steps in mobile sensing assessments and (ii) identify the optimal levels of aggregation for mobile sensing data (e.g. for obtaining measures of intra-individual and inter-individual differences in behavioural patterns).

In addition, with replication at the forefront of many of our conversations, the credibility of predictive analytics used in personality science depends on careful, consistent reporting standards. Although the correct use of machine learning practices provides a built-in defense against many of the issues that have plagued other approaches, concerns about new machine learning specific issues are warranted (e.g. see Stachl et al. 2020 for details on such considerations). For example, sophisticated open-source machine learning frameworks (e.g. AutoML, Create ML, and AutoKeras) largely automate machine learning workflows, making these complex modelling techniques available to even the most naive of users. However, the application of machine learning models can lead to methodological

oversights that have ethical implications (e.g. model biases and feedback loops that impact individuals or groups in disparate ways) and can also lead to unreliability in the predictive models over time (e.g. concept drifts). Such issues have been considered in various research communities[4] and are no less important in personality sensing as a burgeoning area of personality research.

Personality sensing research also needs a set of best practices for the various ethical considerations associated with passive assessment of people's behaviours, thoughts, and feelings and situational contexts. For example, personality sensing researchers should consider the following factors when designing their studies: (i) ways of promoting transparency about data practices (e.g. collection, sharing, and modelling of sensing data), (ii) enabling opt-in features for data collection (e.g. permitting participants to choose which types of data to provide), (iii) providing participants with control of personal data (e.g. allowing participants to redact or delete their data), (iv) incorporating features of self-tracking technologies in the design of their sensing studies (e.g. providing participants with the ability to select what is tracked and receive feedback based on their data to obtain self-insight or behaviour change; for a review, see Kersten-van Dijk, Westerink, Beute, & IJsselsteijn, 2017), and (v) generally adopting approaches that respect individual privacy (e.g. treating informed consent like a process, using sensing software that computes inferences on the device instead of storing raw sensor data, and integrating privacy-by-design principles; Harari, 2020).

With increasing research efforts being directed at the utilization of consumer electronics for sensing research, several security considerations and potential dangers of data breaches must also be taken into account. For example, many sensing studies record data in its raw form, but new legislation (e.g. General Data Protection Regulation in the European Union) encourages stakeholders (e.g. researchers, industry companies, and government organizations) to minimize the amount of personal data collected and processed. Although exceptions for researchers are possible, we suggest that researchers implement privacy-by-design principles in their procedures whenever possible (e.g. Beierle et al., 2018). For example, procedures such as increased data control for participants, on-device data aggregation (e.g. collecting typing behaviour data without recording readable text; Buschek et al., 2018), or distributed, on-device model updates[5] can be used to circumvent the need to collect raw data from the device (McMahan, Moore, Ramage, Hampson, & Arcas, 2016).

---

[3]For more information, see https://lifesensingconsortium.org/ and https://md2k.org/.

[4]For an example of an interdisciplinary workshop on such topics in the health domain, see the Bias in Big Data Workshop from the CONNECT research program at Northwestern University (https://isgmh.northwestern.edu/bias-in-big-data/).

[5]This is typically referred to as 'federated learning', which involves distributed training data on mobile devices and remotely updated models (see McMahan, Moore, Ramage, Hampson, & Arcas, 2017 for additional information).

Towards a behavioural personality science

The usage of digital data records for the automated assessment of personality is currently one of the main goals of personality sensing research. However, taking a more bottom-up approach to devising new ways to quantify individual differences in characteristic patterns of behaving will require researchers to think 'outside the box' of current theoretical frameworks. One of the core new directions we anticipate personality sensing will enable is the pursuit of a more behaviourally focused personality science. By leveraging sensing technologies for behavioural assessment, researchers will be able to model individual differences in real-world behaviour in a way that is unprecedented. For example, smartphone data could be used to derive new behavioural dimensions and classes that describe meaningful variation in behavioural traits (e.g. tendencies to engage in certain behaviours on weekdays vs. weekends) and types (e.g. a person being classified as having a 'working professional' type of lifestyle). Such behavioural dimensions and classes may be entirely different from existing lexical approaches to describing and assessing individual differences.

Thus, personality sensing research could enable the creation of a new behavioural taxonomy derived from sensing data. However, a main challenge of this approach will be the identification of suitable metrics that quantify variation in personality, are predictive of relevant life outcomes, and, ideally, are understandable to humans. These metrics could be identified through a data-driven bottom-up approach but should still relate to known domains of individual differences based on past theoretical work. Once suitable measures are found, their temporal variability and their predictive power will need to be evaluated (see Stachl et al. 2020 for details on how this can be done with machine learning). A second challenge is that the meaning of data-driven metrics may change over time (e.g. popularity of certain smartphone apps for socializing), possibly even more than conventional questionnaire items change in meaning (e.g. 'is talkative'). The creation of such taxonomies will therefore require the continuous relativization of an individual's scores to scores of a relevant norm (e.g. based on year and culture). Furthermore, these scores need to be updated and re-evaluated over time, just as questionnaire items require revision and updates over time.

Initial steps toward this direction have been proposed for the case of text and language data (Boyd & Pennebaker, 2017). However, as personality is thought to be manifested in all types of everyday behaviour, a more holistic approach should prove to be beneficial and generative for future research. To construct a more extensive theory of personality, researchers should consider all possible forms of sensing data available to them when deriving such a taxonomy. In sum, personality sensing will provide a window onto people's unique behavioural patterns in daily life. The insights obtained from this vantage point are sure to move the field towards a more behavioural personality science, taking us one step closer to realizing Theophrastus' vision with modern digital technologies.

FUNDING INFORMATION

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

REFERENCES

Abdullah, S., Murnane, E. L., Matthews, M., Kay, M., Kientz, J. A., Gay, G., & Choudhury, T. (2016). *Cognitive rhythms: Unobtrusive and continuous sensing of alertness using a mobile phone.* Proceedings of the 2016 ACM Internatinal Joint Conference on Pervasive and Ubiquitous Computing, 178–189. https://doi.org/10.1145/2971648.2971712

Abowd, G. D., & Dey, A. K. (1999). Towards a better understanding of context and context-awareness. In *International symposium on handheld and ubiquitous computing*, Springer, Berlin, Heidelberg., 304–307. https://doi.org/10.1007/3-540-48157-5_29

Ai, P., Liu, Y., & Zhao, X. (2019). Big Five personality traits predict daily spatial behavior: Evidence from smartphone data. *Personality and Individual Differences*, *147*, 285–291. https://doi.org/10.1016/j.paid.2019.04.027

Apple iOS Core Motion (2017). https://developer.apple.com/reference/coremotion

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

Beierle, F., Tran, V. T., Allemand, M., Neff, P., Schlee, W., Probst, T., … Zimmermann, J. (2018). Context data categories and privacy model for mobile data collection apps. *Procedia Computer Science*, *134*, 18–25. https://doi.org/10.1016/j.procs.2018.07.139

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

Block, J. (1989). Critique of the Act Frequency approach to personality. *Journal of Personality and Social Psychology*, *56*, 234–245. https://doi.org/10.1037/0022-3514.56.2.234

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

Buschek, D., Bisinger, B., & Alt, F. (2018). *ResearchIME. Proceedings of the 2018 CHI conference on human factors in computing systems–CHI'18*, 1–14. https://doi.org/10.1145/3173574.3173829

Buss, D. M., & Craik, K. H. (1980). The frequency concept of disposition. *Journal of Personality*, *48*, 379–392. https://doi.org/10.1111/j.1467-6494.1980.tb00840.x

Buss, D. M., & Craik, K. H. (1983). The Act Frequency approach to personality. *Psychological Review*, *90*, 713–713. https://doi.org/10.1037/h0021465

Campbell, A. T., Eisenman, S. B., Lane, N. D., Miluzzo, E., Peterson, R. A., Lu, H., … Ahn, G. S. (2008). The rise of people-centric sensing. *IEEE Internet Computing*, *12*, 12–21. https://doi.org/10.1109/MIC.2008.90

Chen, Z., Lin, M., Chen, F., Lane, N., Cardone, G., Wang, R., … Cambell, A. (2013). *Unobtrusive sleep monitoring using smartphones*, (June 2016). https://doi.org/10.4108/pervasivehealth.2013.252148

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

Choudhury, T., & Basu, S. (2005). Modeling conversational dynamics as a mixed-memory Markov process. *Advances in Neural Information Processing Systems*, *17*, 281–288.

Eagle, N., & Pentland, A. S. (2003). *Wearables in the workplace: Sensing interactions at the office*. Proceedings of the Seventh IEEE International Symposium on Wearable Computers. https://doi.org/10.1109/ISWC.2003.1241426

Fast, L. A., & Funder, D. C. (2008). Personality as manifest in word use: Correlations with self-report, acquaintance report, and behavior. *Journal of Personality and Social Psychology*, *94*, 334–346. https://doi.org/10.1037/0022-3514.94.2.334

Fleeson, W. (2001). Toward a structure-and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, *80*, 1011–1027. https://doi.org/10.1037/0022-3514.80.6.1011

Funder, D. C. (2001). Personality. *Annual Review of Psychology*, *52*, 197–221. https://doi.org/10.1146/annurev.psych.52.1.197

Funder, D. C. (2006). Towards a resolution of the personality triad: Persons, situations, and behaviors. *Journal of Research in Personality*, *40*, 21–34. https://doi.org/10.1016/j.jrp.2005.08.003

Funder, D. C. (2009). Naive and obvious questions. *Perspectives on Psychological Science*, *4*, 340–344. https://doi.org/10.1111/j.1745-6924.2009.01135.x

Furr, R. M. (2009). Personality psychology as a truly behavioral science. *European Journal of Personality*, *23*, 369–401. https://doi.org/10.1002/per.724

Ghosh, D. & Vogt, A. (2012). *Outliers: An evaluation of methodologies*. In *Joint statistical metings*, 3455–3460.

Google Activity Recognition API (2017). https://developers.google.com/android/reference/com/google/android/gms/location/ActivityRecognitionApi

Harari, G. M. (2020). A process-oriented approach to respecting privacy in the context of mobile phone tracking. *Current Opinion in Psychology*, *31*, 141–147. https://doi.org/10.1016/j.copsyc.2019.09.007

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

Harari, G. M., Müller, S. R., & Gosling, S. D. (2018). *Naturalistic assessment of situations using mobile sensing methods*. The Oxford Handbook of Psychological Situations.

Harari, G. M., Müller, S. R., Mishra, V., Wang, R., Campbell, A. T., Rentfrow, P. J., & Gosling, S. D. (2017). Students' interest in and compliance with self-tracking methods: Recommendations for incentives based on three smartphone sensing studies. *Social Psychological and Personality Science*, *8*, 479–492. https://doi.org/10.1177/1948550617712033

Harari, G. M., Muller, S. R., Stachl, C., Wang, R., Wang, W., Bühner, M., … Gosling, S. D. (2020). Sensing sociability: Individual differences in young adults' conversation, calling, texting and app use behaviors in daily life. *Journal of Personality and Social Psychology*, *119*(1), 204–228.

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

Harari, G. M., Stachl, C., Muller, S. R., & Gosling, S. D. (in press). *Mobile sensing for studying personality dynamics*. In Handbook of Personaliy Dynamics and Processes.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, *466*(7302), 29–29.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, *2*(1999), 102–138.

Kalimeri, K., Lepri, B., & Pianesi, F. (2013). *Going beyond traits, (December), 27–34*. https://doi.org/10.1145/2522848.2522878

Kersten-van Dijk, E. T., Westerink, J. H. D. M., Beute, F., & IJsselsteijn, W. A. (2017). Personal informatics, self-insight, and behavior change: A critical review of current literature. *Human Computer Interaction*, *32*, 268–296. https://doi.org/10.1080/07370024.2016.1276456

Khwaja, M., Vaid, S. S., Zannone, S., Harari, G. M., Faisal, A. A., & Matic, A. (2019). Modeling personality vs. modeling personalidad: In-the-wild mobile data analysis in five countries suggests cultural impact on personality models. In *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *3*(3), 1–24, https://doi.org/10.1145/3351246

Kroencke, L., Harari, G. M., Katana, M., & Gosling, S. D. (2019). Personality trait predictors and mental well-being correlates of exercise frequency across the academic semester. *Social Science & Medicine*, *236*, 112400. https://doi.org/10.1016/j.socscimed.2019.112400

Lane, N. D., Miluzzo, E., Lu, H., Peebles, D., & Choudhury, T. (2010). *A survey of mobile phone sensing*. IEEE Communications Magazine, (September), 140–150.

Lane, N. D., Mohammod, M., Lin, M., Yang, X., Lu, H., Ali, S., … Campbell, A. (2012). Bewell: A smartphone application to monitor,model and promote wellbeing. In *Proceedings of the 5th international ICSTconference on pervasive computing technologies for healthcare* (p. 23e26) New York, NY: IEEE. https://doi.org/10.4108/icst.pervasivehealth.2011.24616

Lathia, N., Pejovic, V., Rachuri, K. K., Mascolo, C., Musolesi, M., & Rentfrow, P. J. (2013). Smartphones for large-scale behavior change interventions. *IEEE Pervasive Computing*, *12*, 66–73. https://doi.org/10.1109/MPRV.2013.56

Lathia, N., Sandstrom, G. M., Mascolo, C., & Rentfrow, P. J. (2017). Happier people live more active lives: Using smartphones to link happiness and physical activity. *PLoS ONE*, *12*, 1–13. https://doi.org/10.1371/journal.pone.0160589

Likamwa, R., Liu, Y., Lane, N. D., & Zhong, L. (2011). *Can your smartphone infer your mood?* PhoneSense, 1–5.

Lu, H., Pan, W., Lane, N. D., & Choudhury, T. (2009). *SoundSense: Scalable sound sensing for people-centric applications on mobile phones*. Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services, 165–178.

Lu, H., Rabbi, M., Chittaranjan, G. T., Frauendorfer, D., Mast, M. S., Campbell, A. T., … Choudhury, T. (2012). *StressSense: Detecting stress in unconstrained acoustic environments using smartphones*. Ubicomp, 351–360. https://doi.org/10.1145/2370216.2370270

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. y. (2016). *Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017*. Retrieved from http://arxiv.org/abs/1602.05629

McMahan, B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. *In Artificial Intelligence and Statistics*, 1273–1282.

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*, 184–190. https://doi.org/10.1109/EMBC.2016.7590696.Upper

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

Miluzzo, E., Papandrea, M., Lane, N. D., Lu, H., & Campbell, A. T. (2010). *Pocket, bag, hand, etc.—Automatically detecting phone context through discovery*. PhoneSense 2010: Int. Workshop Sensing for App Phones, 21–25.

Mohr, D., Zhang, M., & Schueller, S. M. (2017). *Personal sensing: Understanding mental health using ubiquitous sensors and machine learning.* Ssrn. https://doi.org/10.1146/annurev-clinpsy-032816-044949

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

Montag, C., Błaszkiewicz, K., Sariyska, R., Lachmann, B., Andone, I., Trendafilov, B., … Markowetz, A. (2015). Smartphone usage in the 21st century: Who is active on WhatsApp? *BMC Research Notes, 8,* 4–9. https://doi.org/10.1186/s13104-015-1280-z

De Montjoye, Y., Quoidbach, J., & Robic, F. (2013). *Phone-based metrics,* 48–55.

Moskowitz, D. S., & Zuroff, D. C. (2004). Flux, pulse, and spin: Dynamic additions to the personality lexicon. *Journal of Personality and Social Psychology, 86,* 880–893. https://doi.org/10.1037/0022-3514.86.6.880

Müller, S. R., Harari, G. M., Mehrotra, A., Matz, S., Khambatta, P., Musolesi, M., … Rentfrow, P. J. (2017). *Using human raters to characterize the psychological characteristics of GPS-based places.* Adjunct Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, 157–160. https://doi.org/10.1145/3123024.3123135

Muller, S. R., Peters, H., Matz, S., Wang, W., & Harari, G. M. (in press). Investigating the relationships between mobility behaviors and indicators of subjective well-being using smartphone-based experience sampling and GPS tracking. *European Journal of Personality.*

Murnane, E. L., Abdullah, S., Matthews, M., Kay, M., Kientz, J. A., Choudhury, T., … Cosley, D. (2016). Mobile manifestations of alertness: Connecting biological rhythms with patterns of smartphone app use. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, 2016,* 465–477. https://doi.org/10.1145/2935334.2935383

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science, 349,* aac4716. doi:10.1126/science.aac4716

Ozer, D. J., & Benet-Martínez, V. (2005). Personality and the prediction of consequential outcomes. *Annual Review of Psychology, 57,* 401–421. https://doi.org/10.1146/annurev.psych.57.102904.190127

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In R. C. Robins, & R. F. Kreuger (Eds.), *Handbook of research methods in personality psychology* (pp. 224–239). New York: Guilford.

Perrin, A. (2019). *Digital gap between rural and nonrural Americans persists.* Retrieved from http://www.pewresearch.org/fact-tank/2018/10/19/5-charts-on-global-views-of-china/

Rabbi, M., Ali, S., Choudhury, T., & Berke, E. (2011). Passive and in-situ assessment of mental and physical well-being using mobile sensors. In *Proceedings of the 13th International Conference on Ubiquitous Computing* (pp. 385–394) New York, NY: ACM. https://doi.org/10.1145/2030112.2030164

Radu, V., Tong, C., Bhattacharya, S., Lane, N. D., Mascolo, C., Marina, M. K., & Kawsar, F. (2018). Multimodal deep learning for activity and context recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1,* 1–27. https://doi.org/10.1145/3161174

Ram, N., Yang, X., Cho, M.-J., Brinberg, M., Muirhead, F., Reeves, B., & Robinson, T. N. (2020). Screenomics: A new approach for observing and studying individuals' digital lives. *Journal of Adolescent Research, 35,* 16–50. https://doi.org/10.1177/0743558419883362

Rauthmann, J. F., Sherman, R. A., & Funder, D. C. (2015). Principles of situation research: Towards a better understanding of psychological situations. *European Journal of Personality, 29,* 363–381. https://doi.org/10.1002/per.1994

Rauthmann, J., & Sherman, R. (2018). *Toward a research agenda for the study of situation perceptions: A variance componential framework.* Personality and Social Psychology Review. https://doi.org/10.1177/1088868318765600

Reeves, B., Ram, N., Robinson, T. N., Cummings, J. J., Giles, C. L., Pan, J., Chiatti, A., … Yeykelis, L. (2019). Screenomics: A framework to capture and analyze personal life experiences and the ways that technology shapes them. *Human Computer Interaction,* 1–52. https://doi.org/10.1080/07370024.2019.1578652

Rhodes, R. E., & Smith, N. E. I. (2006). Personality correlates of physical activity: A review and meta-analysis. *British Journal of Sports Medicine, 40,* 958–965. https://doi.org/10.1136/bjsm.2006.028860

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

Rozin, P. (2001). Social psychology and science: Some lessons from Solomon asch. *Personality and Social Psychology Review, 5,* 2–14. https://doi.org/10.1207/S15327957PSPR0501_1

Russell, J. A., & Ward, L. M. (1982). Environmental psychology. *Annual Review of Psychology, 33,* 651–689. https://doi.org/10.1146/annurev.ps.33.020182.003251

Sandstrom, G. M., Lathia, N., Mascolo, C., & Rentfrow, P. J. (2017). Putting mood in context: Using smartphones to examine how people feel in different locations. *Journal of Research in Personality, 69,* 96–101. https://doi.org/10.1016/j.jrp.2016.06.004

Santani, D., Biel, J.-I., Labhart, F., Truong, J., Landolt, S., Kuntsche, E., & Gatica-Perez, D. (2016). *The night is young. Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing–UbiComp'16, (September),* 427–438. https://doi.org/10.1145/2971648.2971713

Schmid Mast, M., Gatica-Perez, D., Frauendorfer, D., Nguyen, L., & Choudhury, T. (2015). Social sensing for psychology. *Current Directions in Psychological Science, 24,* 154–160. https://doi.org/10.1177/0963721414560811

Schoedel, R., Au, Q., Völkel, S. T., Lehmann, F., Becker, D., Bühner, M., … Stachl, C. (2019). Digital footprints of sensation seeking. *Zeitschrift für Psychologie, 226*(4), 232–245. https://doi.org/10.1027/2151-2604/a000342

Schoedel, R., Pargent, F., Au, Q., Völkel, S. T., Schuwerk, T., Bühner, M., & Stachl, C. (2020). To challenge the morning lark and the night owl: Using smartphone sensing data to investigate day-night behavior patterns. *European Journal of Personality, 34,* 733–752. https://doi.org/10.1002/per.2258

Shmueli, G. (2011). To explain or to predict? *Statistical Science, 25,* 289–310. https://doi.org/10.1214/10-sts330

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,* 711–727. https://doi.org/10.1177/0956797619831612

Stachl, C., Au, Q., Schoedel, R., Buschek, D., Völkel, S. T., Oldemeier, M., … Bischl, B. (2019). *Behavioral patterns in smartphone usage predict Big Five personality traits.* PsyArXiv Preprints. https://doi.org/10.31234/osf.io/ks4vd

Stachl, C., Hilbert, S., Au, J. Q., Buschek, D., De Luca, A., Bischl, B., … Bühner, M. (2017). Personality traits predict smartphone usage. *European Journal of Personality, 31,* 701–722. https://doi.org/10.1002/per.2113

Stachl, C., Au, Q., Schoedel, R., Gosling, S. D., Harari, G. M., Buschek, D., … Bühner, M. (2020). Predicting Personality from Patterns of Behavior Collected with Smartphones. *Proceedings of the Natural Academy of Sciences of the United States of America (PNAS).*

Taylor, B. Y. K., & Silver, L. (2019). *Pew Research Center. "Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally"*. Retrieved from: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/

Teso, S., Staiano, J., Lepri, B., Passerini, A., & Pianesi, F. (2013). *Ego‐centric graphlets for personality and affective states recognition*. IEEE International Conference on Social Computing, 874–877. https://doi.org/10.1109/SocialCom.2013.132

Vaizman, Y., Ellis, K., & Lanckriet, G. (2017). Recognizing detailed human context in the wild from smartphones and smartwatches. *IEEE Pervasive Computing*, 16, 62–74. https://doi.org/10.1109/MPRV.2017.3971131

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., … Campbell, A. T. (2014). StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*, 3–14.

Wang, R., Harari, G., Hao, P., Zhou, X., & Campbell, A. T. (2015). *SmartGPA: How smartphones can assess and predict academic performance of college students*. UbiComp 2015—Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 295–306. https://doi.org/10.1145/2750858.2804251

Wang, R., Wang, W., daSilva, A., Huckins, J. F., Kelley, W. M., Heatherton, T. F., & Campbell, A. T. (2018). Tracking depression dynamics in college students using mobile phone and wearable sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2, 1–26. https://doi.org/10.1145/3191775

Wang, W., Harari, G. M., Wang, R., Müller, S. R., Mirjafari, S., Masaba, K., & Campbell, A. T. (2018). Sensing behavioral change over time: Using within-person variability features from mobile sensing to predict personality traits. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2, 141. https://doi.org/10.1145/3264951

Wrzus, C., & Mehl, M. R. (2015). Lab and/or field ? *Measuring Personality Processes and Their Social Consequences*, 271, 250–271.

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

Zeigler-Hill, V., Shackelford, T., Nave, C. S., Feeney, M. G., & Furr, R. M. (2018). *Behavioral observation in the study of personality and individual differences*. The SAGE Handbook of Personality and Individual Differences: Volume I: The Science of Personality and Individual Differences, 317–340. https://doi.org/10.4135/9781526451163.n15